MALE AND FEMALE DERIVED REPRODUCTIVE PROTEINS THAT CONTRIBUTE TO THE REGULATION OF THE LONG-TERM POST-MATING RESPONSE IN FEMALE *DROSOPHILA MELANOGASTER*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Jessica Lynn Sitnik

January 2014

MALE AND FEMALE DERIVED REPRODUCTIVE PROTEINS THAT CONTRIBUTE TO THE REGULATION OF THE LONG-TERM POST-MATING RESPONSE IN FEMALE *DROSOPHILA MELANOGASTER*

Jessica Lynn Sitnik, Ph. D.

Cornell University 2014

In organisms with internal fertilization, seminal fluid proteins (Sfps) are essential for the reproductive success of both sexes. In *Drosophila melanogaster*, Sfps are required to initiate and maintain females' post-mating responses (PMR). The Drosophila PMR include changes in egg laying, receptivity to courting males, and sperm storage. Previous studies have identified functions for only a handful of the 208 Sfps identified to date. The best-characterized Sfp is the "sex peptide" (SP), which is necessary for many of the sustained aspects of the PMR. Five other Sfps (CG9997, CG1656/1652, CG17575, and seminase) and one female protein (sex peptide receptor (SPR)) were known to be essential for SP's function to persist (the "long term response network" (LTR network)). How the LTR is modulated, however, is not understood.

Here, I present studies that identified and characterized the functions of new male- and female-contributors to the PMR and LTR; these new proteins were identified by using or integrating evolutionary, molecular, mutational, and targeted gene approaches. First, I report that the *Drosophila* orthologs of mammalian *Neprilysin* proteins play a conserved role in both male and female fertility. Further, I show that in female flies the insect-specific *Nep2* is important to regulate long term egg-laying, sperm storage, and sperm utilization making it an attractive target for regulation by Sfps. Second, I describe the evolutionary relationships and functions of a

family of three gene duplicates, one of which encodes an Sfp that appears to have arisen by neofunctionalization and subsequent co-option of the duplicate of a female specific gene. I demonstrated that the female-expressed CG32834 is important for short term egg-laying and that this gene and the other female-expressed family member, CG9897, regulate long term receptivity whereas the Sfp, CG32833, regulates overall egg-laying. Third, I used RNAi to show LTR-network function for the Sfp *Intrepid* (CG12558), whose sequence shows evolutionary rate covariation with previously known LTR network proteins. Finally, to determine the cellular source of LTR network proteins, I examined the roles of the two cell types in male flies' accessory glands (the source of network proteins). I characterized the reproductive phenotypes of *iab-6*^{cocu} males, which are deleted for an enhancer in the Hox gene *Abdominal-B (Abd-B)*; the accessory glands of these males lack large vacuole filled secondary cells. I found that products of the secondary cells are required for long term changes in egg laying and receptivity in post-mated females, and are influential during sperm competition. Further, using a secondary cell specific driver derived from the *iab-6* enhancer I determined that the LTR network proteins CG1656/CG1652 and CG17575 are produced specifically in the secondary cells. Using RNA-seq data to identify genes down-regulated in *iab-6*^{cocu} males combined with secondary cell specific RNAi my collaborators and I identified eight additional genes whose expression in the secondary cells is required for the LTR. Only one of these genes, CG3349, encodes a known transferred Sfp, suggesting that the *iab-6*^{cocu} mutation may not primarily affect Sfps directly but instead might work through disrupting other cellular functions. Together, my results significantly increase our knowledge of the actions and origins of male and female molecular regulators of post-mating responses, as well as of the integrated roles of the cell types that produce the male regulators.

BIOGRAPHICAL SKETCH

Jessica Lynn Sitnik was born and raised in Virginia. She is the only child of Peter and Patricia Sitnik. Jessica's interest in the natural world started early, shifting between paleontology in her early years to marine biology, ecology, and genetics as time passed. She has always been asking questions, sometimes to the detriment of her parent's sanity, like that one time she kept 20 crayfish in individual tanks to test whether different levels of exposure to malathion increased their shell pliability thus weakening their defense against predation and increasing the likelihood that the poison would spread up the food chain (the results were sadly inconclusive in the end). During high school Jessica interned at the Smithsonian National Museum of Natural History with Dr. James G. Mead , where she helped his students extract teeth from rough-toothed whales (*Steno bredanensis)* for analysis, moved specimens between freezers, and in general learned about what it's like to be a researcher. After high school, Jessica attended the College of William and Mary where she received a B.S in Biology and completed the requirements for a minor in studio art. During her time at William and Mary Jessica worked in the lab of Dr. Gregory Capelli investigating the effects of sex ratio on mate guarding behavior in the freshwater amphipod *Gammarus pseudolimnaeus*. She also was part of the REU program at the Keck Environmental Field Laborotory where, under the guidance of both Dr. Capelli and Dr. Randolph Chambers, she investigated the effect of storm water retention ponds on macro invertebrate community structure as a means to assess their effectiveness in protecting stream systems. Jessica decided that although she loved marine biology and ecology she wanted a finer lens through which she could address the questions that interested her. After graduation she moved to Ithaca, New York to begin graduate school at Cornell University in Genetics and Development. Once there she joined the lab of Dr. Mariana Wolfner where she studied a different kind of

environmental influence on behavior, the chemical signals contained in the seminal fluid of males and transferred to females during mating, using the model system *Drosophila melanogaster*.

This thesis is dedicated to all those who believed in me when I could not believe in myself. If not for your kind faces, gestures, and words I would not be where I am today. You know who you are...and I will never forget.

ACKNOWLEGEMENTS

First I would like to thank my advisor and mentor, Mariana Wolfner. Mariana is the kind of mentor I always wanted but didn't know I needed. When I first came to Cornell I really struggled with graduate school and my own personal insecurities. I was floundering and where one person could have seen a waste of time and energy she saw potential and someone who just needed the right support. I am very grateful that she took a chance on me. I hope that over the years I have lived up to what she saw in me and that I continue to do so. In addition, Mariana is a great motivator. She has a way of reframing things that can make something you don't want to do or aren't sold on seem exciting and suddenly of pivotal importance. It's like she had a plan all along that only she knew. Mariana was also understanding of my desire to improve my teaching and mentoring ability and was wonderful to talk with about all things science and education related. At the same time she was good about helping me stay focused, something I desperately needed at times. She's also remarkably funny in her own right, especially that twinkle in her eye she gets when she thinks she's being clever (which usually she is!) It has been a pleasure to be part of her lab.

I would also like to thank my committee members, Dan Barbash and Eric Alani. I have a great admiration and respect for you both and I truly value our discussions and your suggestions over the years (about science and otherwise). I'd like to thank my undergraduate advisor Gregory Capelli for giving me the opportunity to work in his lab, helping me apply for the REU program, and vouching for me when Cornell called to make sure I was a good fit for them. Greg was a kind soul who was a pleasure to talk with about the nature of people and the lives of animals. I valued his friendship and wish he was here to see what his efforts helped me to accomplish. I thank James Mead for allowing a lowly high school student to pad around the

Smithsonian. I learned so much and I came away with some wonderful, albeit somewhat gross, stories. Like the time the walk in freezers full of whale carcasses broke. We had to chisel partially thawed and refrozen specimens off the shelves to pass, assembly line style, into bins for relocation. I smelled so bad afterwards that no one would sit near me on the metro ride home, yet somehow I still ended up in science. I also thank the excellent teachers I've had over the years, especially my middle school science teacher Timothy Riegner and my AP Biology teacher Malcom Leinwohl. They both had a love for knowledge and learning that inspired me to seek answers and ask questions. It was Mr. Leinwohl's idea, to require his AP Biology students to intern or shadow with someone working in Biology, that brought me to James Mead's door.

I thank all the members of the Wolfner Lab. Between our scientific discussions (or not so scientific) and the way everyone supports each other you've made Cornell a great place to work. I'd especially like to thank Norene Buehner, without whom the lab would fall to pieces, for all of her support and friendship along the way, Frank Avila, who was my go to guy in times of academic crisis (or new and interesting movies/music), and Geoff Findlay, for his sense of humor and thoughtful analysis which helped me evaluate many a failed experiment. I also thank my friends outside the lab for reminding me that there is always still time to do something fun now and again and that growing up doesn't mean silliness is no longer allowed.

I would also like to thank my family, especially my parents. Thank you for instilling in me a love of nature and the unknown coupled with a need for understanding. Without your support and love I would not be where I am today. Also, thanks for trying to understand what I'm working on and asking me questions. Your interest and curiosity means the world to me. I could thank you for all the other parent-y things you do here too, but I think I have to include some space for a thesis somewhere.

And perhaps most importantly, I thank my loving husband Andy Jefferson. Without him I would never have made it through undergrad, much less graduate school. Thank you for all the late night talks about human nature, learning, and science, for cooking when I was too tired, for cleaning when I was too lazy, for being silly when I was too serious, and for going out and finding interesting people to bring into our lives. I may ground you as you float around dreaming of a better tomorrow, but you keep me from letting worry, fear, and despair take over a bright today. I can never thank you enough for all the love and support you've given me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Often when we think of sexual reproduction our focus is on fertilization, the union of egg and sperm. While this process is important, it certainly isn't the only step essential for reproductive success. Along with sperm, males transfer proteins in their ejaculate. These proteins, are collectively referred to as seminal fluid proteins (Sfps). Sfps greatly impact the sperm transferred with them and they also interact with the female, locally in the reproductive tract and more globally in the central nervous system, resulting in changes in female gene expression, behavior, and physiology (reviewed in [1]).

Sfps are found in all known sexually reproducing organisms, from insects to mammals, and underlie a diverse set of changes in the female after mating. In llamas, seminal fluid contains an ovulation-inducing factor, recently identified to be β nerve growth factor (β-NGF) [2] (reviewed in [3]). Exposure to β-NGF from the male results in an increase in the release of lutenizing hormone (LH) from the pituitary of the mated female, triggering ovulation. In the honeybee, *Apis mellifera L*., seminal fluid proteins along with secretions from the sperm storage organs of the female play a role in the survival of sperm in storage [4,5]. And in *Drosophila melanogaster*, females who receive the Sfp sex peptide (SP) from their mates experience an increase in egg-laying and are less likely to mate with subsequent males [6,7]. The collective changes that are observed in females after mating are referred to as the female post-mating response (PMR).

Despite what we know about Sfps and their importance for regulating the female PMR, most of the Sfps identified to date have unknown functions. In addition, the mechanisms behind

how Sfps influence the female and interact with each other are still largely unclear and where they have been pieced together partially (such as in the case of β-NGF or SP) gaps still remain. Here I describe what is known about the regulation of the PMR in one of the best understood model systems, *Drosophila melanogaster*, with a focus on the sustained aspects of the PMR often referred to as the long term response (LTR). I discuss *Drosophila* reproductive proteins that have homologs in other taxa, and that may contribute to the regulation of the LTR in *Drosophila*. I also examine methods to identify Sfps with functions in the PMR both by predictions based on molecular evolution and by targeted gene approaches. Lastly I discuss the structure and development of the male accessory gland, which contributes the vast majority of Sfps to the seminal fluid.

The female post-mating response (PMR) in Drosophila melanogaster[1]

The female PMR in *Drosophila melanogaster* is robust and well characterized (reviewed in [1]). Combined with the myriad of genetic and molecular tools available for this animal, the PMR offers an ideal model for studying the role of male derived Sfps in regulating female behavior. For example, after mating, *Drosophila* females increase their rates of ovulation , egg laying, food intake, excretion, and siesta sleep [8]. Sperm from their mates are stored in specialized sperm storage organs (depicted in Figure 1.1); anatomical changes of the uterine shape occur beginning at the onset of mating that may help to facilitate this process [9,10]. For

[1] A small portion of this section on the female PMR has been modified from text that I contributed to a review published in Advances in Genetics in 2009 as "Laura K. Sirot, Brooke A. LaFlamme, Jessica L. Sitnik, C. Dustin Rubinstein, Frank W. Avila, Clement Y. Chow, Mariana F. Wolfner, "Chapter 2 - Molecular Social Interactions: Drosophila melanogaster Seminal Fluid Proteins as a Case Study" The full text of that section can be found in APPENDIX A. The figure of female rejection (Figure 1.2) is also from that paper. Reprinted with permission.

several days, mated females are less likely to accept suitors, actively fleeing or kicking any persistent male (Figure 1.2). In addition to these more obvious changes, the female undergoes changes in transcription including increases in expression of several known anti-microbial peptide genes within hours after mating [11,12,13,14].

While it is possible that these changes in behavior, physiology, and gene expression could be brought about by either the behavioral act of mating or the transfer of sperm, males that mate but do not produce sperm still elicit post-mating responses in their partners [15] demonstrating that non-sperm components of the seminal fluid must be involved [16]. In fact, males that transfer sperm but do not transfer Sfps produced in the accessory glands (a major site of Sfp synthesis) fail to elicit most post-mating responses [17,18,19]. It is known that the ejaculatory duct and ejaculatory bulb also produce secreted proteins that enter the seminal fluid. Some of these proteins are necessary for post-mating responses as well, possibly through their involvement with the production of the mating plug [20,21,22].

The PMR itself is usually divided into two periods that I will refer to as the short term and long term responses. The short term response (STR) consists of changes that occur in the first 24 hours after mating. It does not require the transfer of sperm. Short term changes include but are not limited to an increase in ovulation rate, the initial spike in egg- laying that occurs in the first 24 hours after mating, increased juvenile hormone production [23], dampening of female receptivity to courting males, increased feeding and excretion, and changes in transcription and siesta sleep. The best characterized Sfp that regulates an STR is the prohormone Ovulin, which is involved in increasing ovulation in the first 24h after mating [8,24,25]. After being transferred to the female ovulin localizes to the base of the ovary [24] and also enters the hemolymph. Ovulin is thought to act through the central nervous system [26], possibly by stimulating

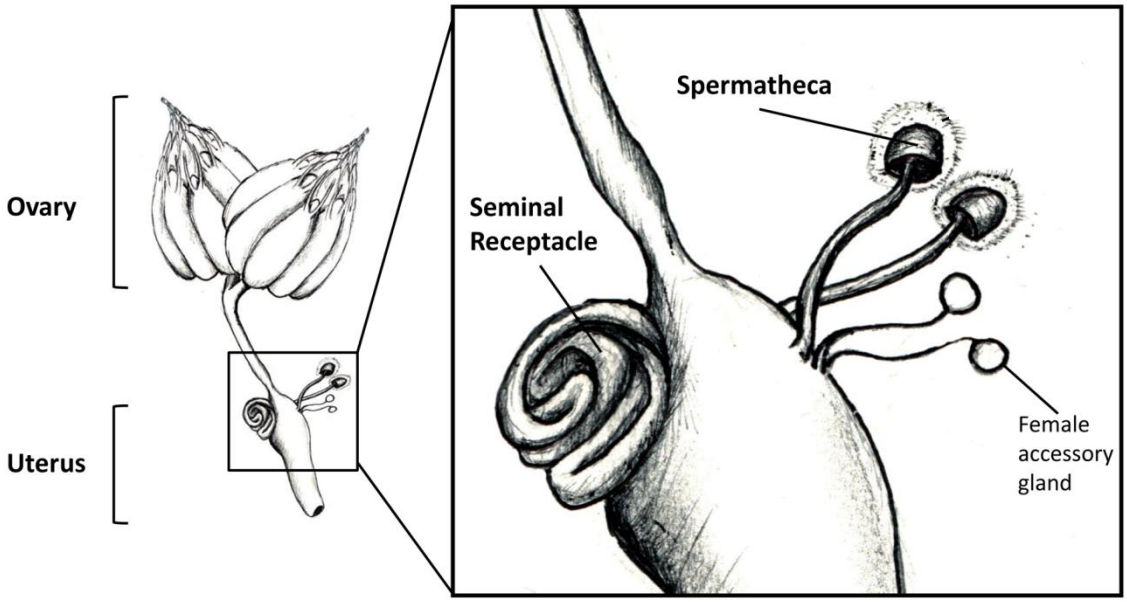


Figure 1.1 The *Drosophila melanogaster* female reproductive tract.

The ovaries produce oocytes that pass down the common oviduct into the uterus before they are deposited during egg-laying. Sperm from the male is stored in specialized organs called the seminal receptacle (SR) and the spermathecae (ST), which are shown magnified in the inset. The parovaria, or female accessory glands, are also depicted but will not be discussed in this thesis. This figure is a slightly modified version of a drawing by J. L. Sitnik and was previously published in PLoS Biology. doi:10.1371/journal.pbio.1001191.g001

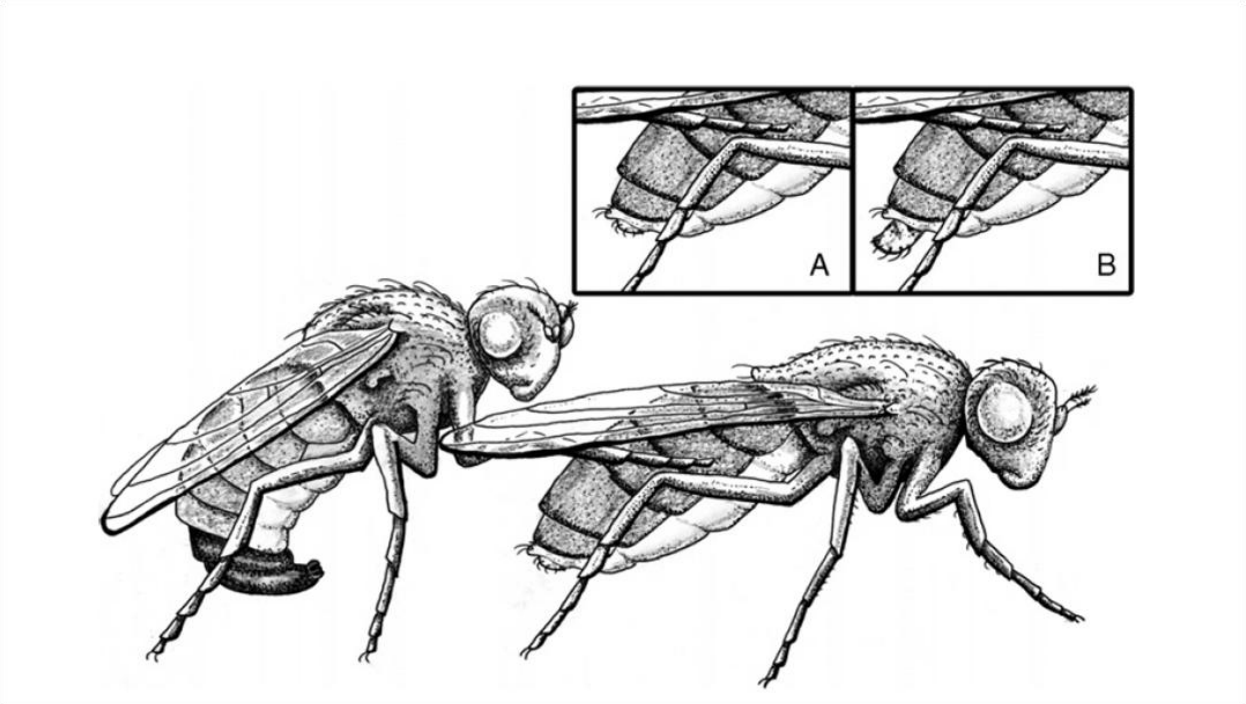


Figure 1.2 Rejection by ovipositor extension in female *Drosophila melanogaster*.

Mating attempt by male and (insets) position of female abdomen when she is receptive (A) and unreceptive (B; ovipositor extruded) to the mating attempt. Drawing by J. Sitnik. Originally published in Sirot LK, LaFlamme BA, Sitnik JL, Rubinstein CD, Avila FW, et al. (2009) Molecular social interactions: *Drosophila melanogaster* seminal fluid proteins as a case study. Adv Genet 68: 23-56.

octopamanergic neurons [27], though a receptor for ovulin has yet to be identified.

The long term response (LTR) can persist up to 14 days. It requires the transfer and storage of sperm [7,28]. In some ways the LTR can be thought of as the processes that maintain a subset of the changes initiated during the STR. However the LTR also includes Sfp influence over stored sperm and their release [29]. Originally the LTR was thought to occur as a direct result of sperm storage, and was termed the "sperm effect" [30,31], However, it was later found to be due to the transfer, storage, and subsequent cleavage of the sex peptide (SP) [6,7,28]. SP is a 36aa-long peptide that affects female response to male courtship, oogenesis, ovulation, immune response, and juvenile hormone production [6,7,23,32]. Loss of SP transfer severely dampens these aspects of the PMR in the STR and ablates them in the LTR, suggesting that SP is necessary for both full initiation of the PMR and its maintenance. Since the main focus of this dissertation is on processes that contribute to the initiation and maintenance of the LTR, the following sections focus on what is known about this part of the PMR.

A network of Sfps is essential for the LTR

In order for SP to initiate its characteristic changes on the PMR it must first be transferred to females during mating. Failure to transfer SP results in some reduced aspects of the STR (full increase in egg-laying and receptivity suppression do not occur) and loss of the LTR. Once inside the female, SP binds to sperm and enters the seminal receptacle where it remains in storage for upwards of 10 days. Over time the C-terminal portion of SP is cleaved off the surface of sperm [28] enabling it to interact with its receptor, the G-protein-coupled-receptor sex peptide receptor (SPR), both in the reproductive tract and in the central nervous system [33].

This gradual release of the C-terminal part of SP is what is thought to allow for the prolonged effects referred to as the LTR. If SP fails to be stored normally either through a failure to bind to sperm, a failure in sperm entry into or retention in storage, or a failure in cleavage off the surface of sperm then the STR occurs normally but cannot be maintained, resulting in a loss of the LTR .

While ectopic expression of SP in females suggests that SP is sufficient to cause many aspects of the PMR [34], the process through which SP is able to bind to sperm, allowing it enter storage and subsequently influence the female CNS long term, requires many additional Sfps acting in a network ([35] and Chapter 4). Candidate-based approaches using RNAi identified 5 Sfps (CG9997, CG1656/CG1652, CG17575, and Seminase) that are essential for the proper storage of SP [35,36,37]. Reduction in the expression of any one of these Sfps results in loss of the LTR (as measured through changes in the number of eggs laid, receptivity to courting males, and stored sperm for mates of RNAi males compared to controls) but unlike loss of SP itself does not impact the STR. Knockdown of individual genes combined with Western blotting using antibodies specific to these LTR proteins allowed them to be placed in a pathway [35,37]. A general outline of the network from male to female can be seen in Figure 1.3 and is described below.

The Sfp CG9997 is a serine-type endopeptidase which is required for the transfer of the lectins CG1656 and CG1652. After transfer to the female, CG1656/CG1652 inhibit the cleavage of CG9997, allowing some of the full-length protease to remain inside the female. The purpose of the cleavage of CG9997, or whether this inhibition of cleavage is necessary for SP storage is unclear. The protease Seminase is necessary for the accumulation of all of the known LTR pathway proteins in the female seminal receptacle but not for their transfer to the female, whereas the cysteine-rich secretory protein (CRISP) CG17575 is only known to be necessary for

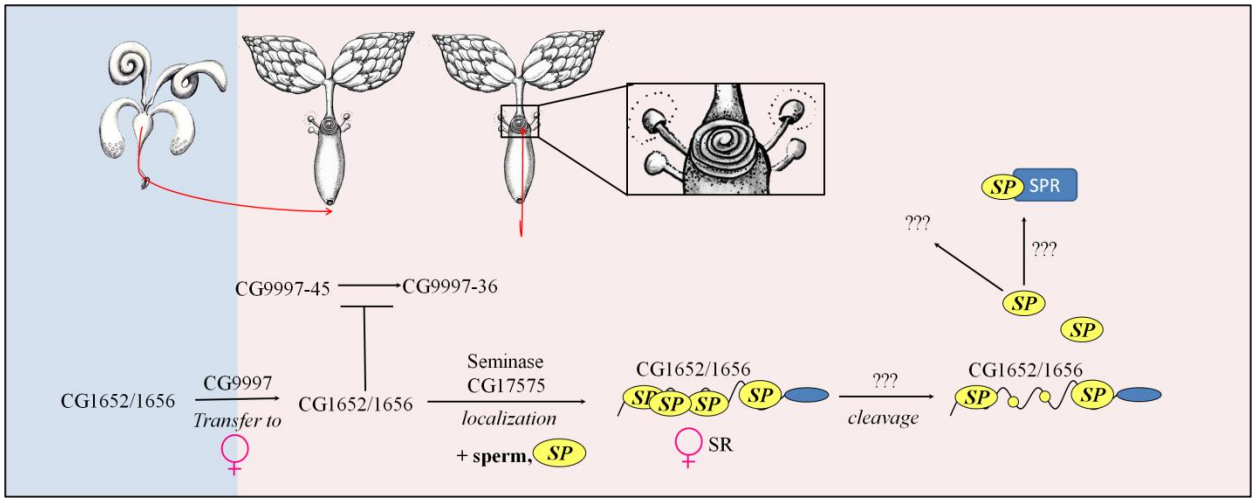


Figure 1.3 A Network of proteins are required for the LTR

Males transfer the seminal fluid proteins CG9997, CG1656, CG1652, CG17575, Seminase, and SP to their mates. CG1656 and CG1652 require CG9997 to be transferred to the female. Once there, CG9997 is stabilized by CG1656 and CG1652, All 6 Sfps localize to the female seminal receptacle (SR), this localization requires the presence of Seminase and CG17575. Either before or during the storage process, SP and CG1656 bind to sperm. After storage, SP is cleaved off the surface of sperm and can then interact with its receptor SPR. The reproductive tracts illustrate the step and location in which events occur. Red arrows indicate the movement of Sfps, first as they are transferred from the male to the female and then as they move into the sperm storage organs. The inset depicts the upper female reproductive tract, including the coiled seminal receptacle where SP is stored.

the accumulation of CG1656, CG1652, and SP in the SR. So far CG1656 and SP are the only pathway constituents that are known to bind directly to sperm and CG1656 requires the presence of CG17575 for this binding to happen. There is currently no evidence that these proteins interact directly with each other or SP, suggesting that there are still gaps in our understanding of how SP binds to sperm and enters the SR for storage. Likewise, the protease responsible for releasing the c-terminus of SP from the sperm remains unidentified.

Female contribution to the LTR

Once the C-terminal portion of SP is released from the sperm [33,38] it can interact with its receptor SPR. SPR is expressed broadly in the central nervous system and the female reproductive tract [39]. More specifically its function in the PMR has been implicated in a small subset of sensory neurons that innervate the female uterus and oviduct. These neurons also express the gene *fruitless(fru),* a DNA-binding transcription factor that is broadly implicated in sex-specific behaviors [40,41,42,43], and *pickpocket (ppk),* which is a subunit of an amiloride-sensitive sodium channel that is a marker for mechanosensory neurons [42]. SPR expression in *fru+ppk+* neurons is necessary for the behavioral changes seen in females after mating and likely works to silence neuronal transmission from these neurons via inhibition of protein kinase A (PKA) [39]. A subset of these *ppk+fru+* neurons also express the sex determination gene *doublesex (dsx).* These *dsx+ppk+fru+* neurons are thought to connect to the *dsx* circuit in the central nervous system and thoracic ganglion downstream of the SP/SPR interaction, however how this process occurs or regulates the PMR is unknown [40]. While the broad picture of which neurons are needed is becoming clear, there are still gaps in our understanding of how

these neurons ultimately control the PMR and how SP and SPR are able to influence the circuit. In Chapter 4, I describe an evolutionary method that discovered additional genes that control the female side of the LTR.

Conserved classes and families of reproductive proteins, including Sfps

Studying the *Drosophila melanogaster* PMR offers insight into how external factors, like seminal fluid from the male, can influence female behavior. However, it is important to keep in mind that many reproductive proteins are rapidly evolving [44,45,46] and Sfps are no exception [45,47,48,49,50]. While the individual sequences of these Sfps are not widely conserved across taxa they do fall into several conserved protein classes that are found in the seminal fluid of most animals studied to date. These classes include proteases, protease inhibitors, acid lipases, cysteine rich secretory proteins (CRISPs), and lectins [36,51]. Protease cascades involving transferred Sfps, in particular, are important in both human and Drosophila reproduction [52]. And, while the specific interactors may be different the general mode of transport/action for some Sfps may also be conserved. For example, Sfps do not have to perform their function locally in the female reproductive tract. Some Sfps, such as β-NGF in mammals [2,3] and SP and Ovulin in flies [33,38,53], are capable of entering circulation and interacting more globally with the female central nervous system. Additionally despite major differences in the regulation of reproduction, some Sfps are found in multiple species. For example, β-NGF is produced in the seminal fluid of all mammals that have been tested to date, including cows, horses, and humans [3,54,55]. Despite differences in ovulation control, β-NGF from spontaneously ovulating species is still capable of inducing ovulation in llamas, suggesting that the features required for this

function are present in these species. This finding may offer insights into the evolution of ovulation control in the female [56].

Conserved male and female reproductive proteins are also of interest, as they potentially provide common ground for Sfps to act upon. One such conserved protein family is the M13 class of neutral endopeptidases often referred to as Neprilysins, which are the subject of Chapter 2 of this thesis. These proteases are membrane-bound zinc-metalloproteases and are involved in the processing of neuropeptides and peptide hormones [reviewed in 57,58,59]. They are important in diverse physiological systems and have been implicated in various diseases including cardiovascular disease [60,61], Alzheimer's disease [62,63], inflammatory disorders [64], and cancer [65,66]. In addition to their role in disease, NEPs are essential for development and reproduction in mammals, but their role there is poorly understood. For example, loss of *NL1*, the mouse ortholog of human *Neprilysin 2*, in males results in a reduction in liter size. However, this change in litter size is not due to defects in spermatogenesis, sperm number, or sperm mobility.

Though Neps have several substrates, the best known are the tachykinins (TKs) a family of related peptides that includes substance P, neurokinin A, and neurokinin B (reviewed in [67,68]. In mammals, TKs are generally present in the central nervous system, where they are considered excitatory neurotransmitters, but they are also found in non-neuronal organs, such as the uterus and placenta of rats and humans [69,70,71]. In female rats and mice, degradation of tachykinins by Neprilysin (NEP) in the uterus is essential for controlling uterine contractions and an inability to degrade tachykinins is associated with a reduction in litter size [72,73].

Neprilysins have been identified among the Sfps produced by the honeybee *Apis*

mellifera [74,75]. Another enzyme with similar substrate affinity to Neps [76], angiotensin-converting enzyme protein (ACE, aka ANCE in flies), is important for spermatogenesis in *Drosophila melanogaster* and present in the secondary cells of the male accessory gland [77,78]. Further, in *Drosophila melanogaster* 24 NEP-like genes have been identified, most of these genes are actively transcribed [39,59,79], and many are expressed in the reproductive tract of either sex and the central nervous system (see Chapter 2). In addition, the *Drosophila* genome encodes TKs and TK receptors[80] that are expressed in brain and thoracic-abdominal ganglion [39]. Together, *Neps* and TKs may contribute to contractions of the uterus, sperm storage organs, and ovary in female *Drosophila* making them excellent targets for Sfps like Acp36DE, which regulates conformational changes in the uterus after mating [10]. Chapter 2 discusses the importance of 5 of these Nep genes in *Drosophila melanogaster* reproduction.

Identification of Sfps that influence the PMR

At this time, 208 proteins are known to be present in the reproductive secretory glands of male *D. melanogaster;* over 90 of these proteins have been confirmed to be transferred to the female along with sperm [36,39,47,81,82,83] (see APPENDIX B for a compiled Table of known and transferred Sfps). Of these, only a handful have been tested for, much less implicated in, regulating any aspects of the PMR. Still, this is far more than the number of molecules known in the female to impact the LTR; in females only SPR is known to interact directly with SP. Other female proteins have been implicated in the PMR, such as the octopamine receptor OAMB [84], however none of these proteins have been shown to interact directly with Sfps. Thus, identifying the function of individual Sfps in the PMR and their interactors in the female is key to expanding our understanding of this process.

Previous studies have largely used candidate gene approaches to investigate the function of individual Sfps, selecting candidates based on similarity to classes of proteins known to be important in other organisms [36,37], those whose expression is specific to or enriched in the accessory glands [8,53,85,86,87], or those whose proteins were able to be isolated and identified from accessory gland extracts [6,88,89,90,91]. These studies are responsible for most of what is known about Sfps to date. However, due to the complexity of the PMR and the difficulty of many PMR associated assays, most of the genes tested in these studies were only assessed for a specific subset of female behaviors. This results in missing out on other potential functions for these genes. Alternative methods for candidate selection, aimed at choosing candidates that are more likely to be associated with specific aspects of the PMR, are needed. Likewise, ways of identifying Sfps that are more likely to interact with female proteins directly would greatly increase our chances of also detecting those female interactors. Using evolutionary methods to identify candidate genes we can get at both of these issues by asking, "Where might Sfp genes come from and how have they changed over time?"

Gene duplication and signatures of co-evolution as a means to detect PMR genes

Evolutionary studies have suggested that positive selection and tandem gene duplication can drive the evolution of seminal fluid proteins between species and, in fact, many tandem gene duplicates exist among the Sfps encoded in the *Drosophila melanogaster* genome[47]. When gene duplication occurs it can allow for the evolution of novel proteins through modification of the original gene or its paralog. This includes changes in expression pattern, which allows for the co-option of genes originally expressed in other tissues, organs, or even sexes. If we consider that some Sfps need to interact directly with female proteins, it seems plausible that one way for

this to occur is by duplication and modification of a gene that is expressed in the female and already has a function in female.

Secretory cells of the sperm storage organs in the *Drosophila melanogaster* female, particularly the secretory cells of the spermathecae, secrete proteins that are necessary for egg-laying, ovulation, and sperm storage [92,93,94,95]. Gene duplication of these proteins produces prime candidates for co-option by the male because they already encode secreted proteins and some perform a function that is advantageous for the male to control. In this context, one Sfp is of particular interest, CG32833. This gene has two tightly linked paralogs, CG9897 and CG32834, that are both expressed in the female spermatheca. All three genes are predicted to encode secreted proteins containing serine-type proteases domains. The function of these genes in the female PMR was unknown and investigating the role of both male and female proteins may shed light on how Sfps evolve and interact with the female protein environment. Chapter 3 investigates the importance of these genes in regulating two aspects of the PMR; egg-laying and female receptivity to remating.

Another method for identifying genes whose products interact with known PMR associated Sfps, either as part of a pathway or in the female, involves looking at signatures of molecular co-evolution. One possible way of bioinformatically inferring functional relationships is through Evolutionary Rate Covariation (ERC) [96]. This method relies on the assumption that functionally related proteins experience similar evolutionary pressures that act on the network as a whole rather than individual proteins..ERC and related methods are often used to study proteins that are known to interact either physically or functionally, but it can be used to predict proteins in a network [97]. Chapter 4 uses the predictive capacity of ERC to identify new members of the LTR pathway in both sexes.

The Drosophila male accessory gland is composed of two distinct secretory cell types

While other secretory tissues, such as the ejaculatory duct and ejaculatory bulb, contribute to the pool of Sfps [20,98] in *Drosophila melanogaster*, the male accessory gland (AG) is the main source of these proteins [82] (Figure 1.4). Further, loss of the gland through a mutation in the Pax gene *paired* is sufficient to result in a loss of the PMR [19]. The accessory gland consists of two lobes, each of which is composed of a monolayer of secretory cells that can be divided into two morphologically distinct cell types. In *Drosophila melanogaster*, flat, polygonally shaped "main cells" make up 96% of each lobe, whereas the remaining 4% of the cells are large, spherical, vacuole filled "secondary cells"[99,100]. These secondary cells are located at the distal tip of the gland in *Drosophila melanogaster* and most other *Drosophila* species investigated to date. Previous studies using enhancer trapping have shown that these two cells types are both morphologically and biochemically distinct [100,101]. Tools to target the secondary cells were elusive, but targeted cell ablation of the main cells [16] results in a loss of both the STR and LTR.

The state of the AG is not unique. Other secretory organs are composed of multiple cell types. Perhaps the most well known example is the islets of Langerhans in the pancreas. Here, 5 cell types (Alpha, Beta, Delta, PP, and Epsilon) work together, three through paracrine feedback, to regulate glucose metabolism (see figure 1.5))[102]. Reproductive secretory tissues, like the epididymis and the prostate in mammals, are also composed of multiple populations of cells. In the epididymis (reviewed in [103]), principal cells make up ~80% of the tubule and are responsible for secreting the majority of the proteins found in the lumen. The rarer basal cells can regulate electrolyte transport in the principal cells by releasing the prostaglandin PGE2 [104]. In addition, apical, narrow, and clear cells of the epididymis secrete protons, aiding in

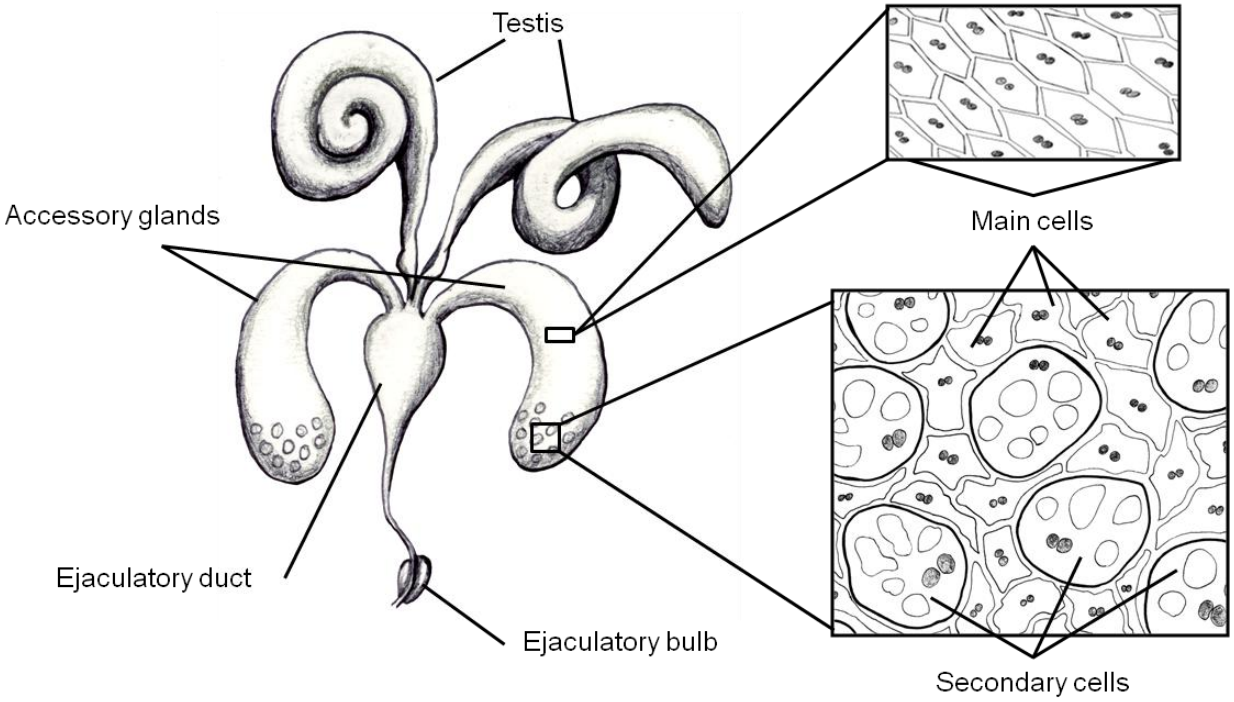


Figure 1.4 The male accessory gland in *Drosophila melanogaster*

A schematic of the male reproductive tract containing the testis, accessory glands, ejaculatory duct, and bulb. The magnified insets show the general relationship between the main and secondary cells, with smaller polygonally shaped main cells tightly packed in the proximal portion of each gland and the larger secondary cells interspersed between the main cells at the distal tip. Both cell types are binucleate. This drawing by J.L.S. was modified from a version originally published in PLoS Genetics, doi:10.1371/journal.pgen.1003395. The full version can be seen in Chapter 5.

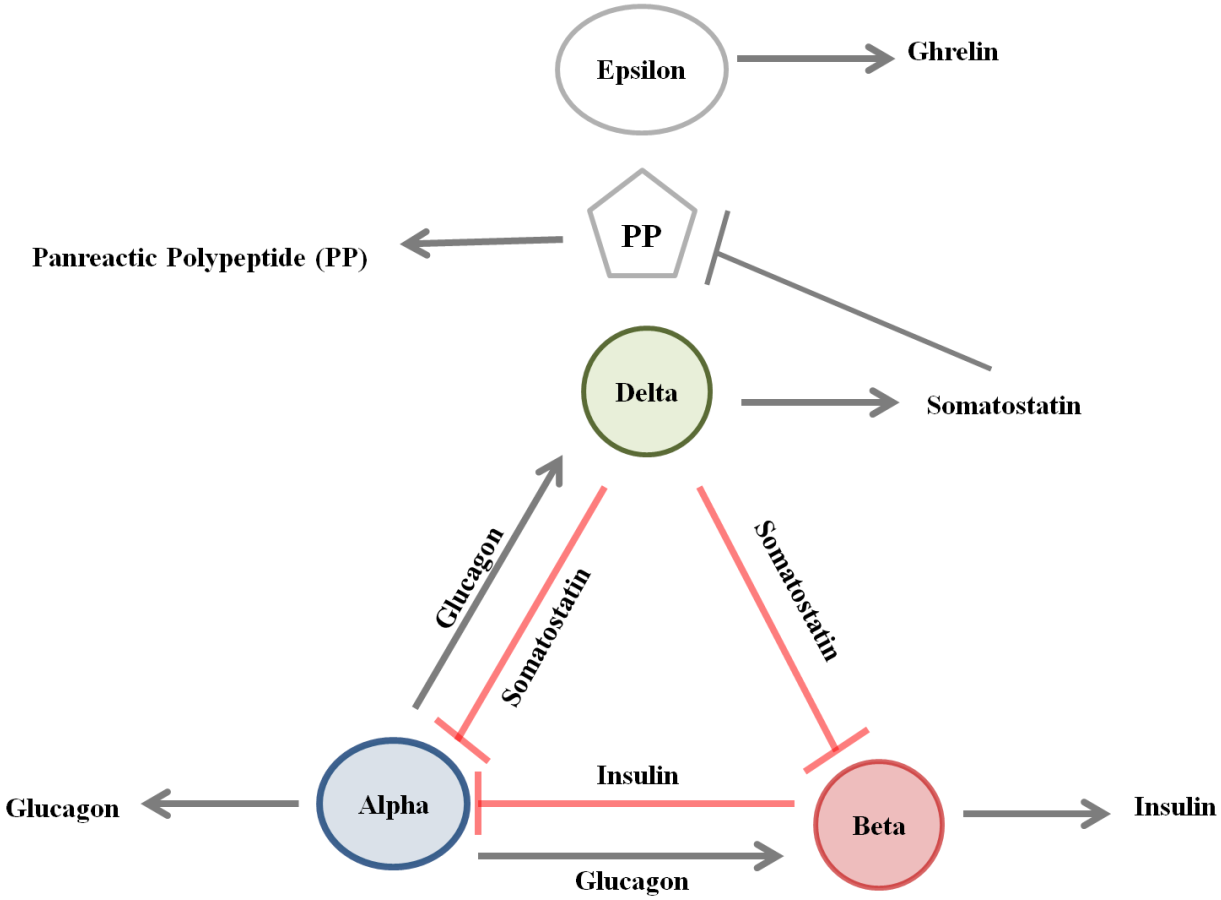


Figure 1.5 The 5 cell types of the islets of Langerhans

The islets of Langerhans are composed of five secretory cell types (Alpha, Beta, Delta, PP, and Epsilon) that secrete hormones directly into the blood. Three of these cell types are involved in a paracrine feedback loop. Alpha cells secrete glucagon, which raises blood glucose levels by stimulating the conversion of glycogen in the liver to glucose. Glucagon stimulates Delta cells to secrete somatostatin, an inhibitory hormone that reduces the secretion of other hormones. The increase in blood glucose caused by glucagon production results in release of insulin by the Beta cells, which also inhibits the production of glucagon in the Alpha cells. The PP cells secrete pancreatic polypeptide (PP), which functions to regulate other pancreatic secretions. PP is also inhibited by somatostatin. Epsilon cells in the pancreas secrete ghrelin, a hormone that is important for regulating hunger, allowing it to indirectly affect blood glucose levels.

acidification of the lumen [105,106]. In rats, the principal cells located in the intermediate zone of the epididymis contain large endocytic vacuoles[107]. Given what is known in these secretory organs, the presence of two cell types in the accessory gland is intriguing and points to the possibility that each performs a unique but coordinated role in regulating the PMR. Recently two homeodomain genes, *defective proventriculous (dve)* [108] and *abdominal-B (Abd-B)* [109] were found to be essential for the normal development and maturation of the secondary cells. In Chapter 5 I discuss the importance of the secondary cells in regulating the PMR, and more specifically the LTR, using a mutant allele of *Abd-B, iab-6*^{cou}.

Dissertation Outline

My studies span several aspects of thinking about the female PMR, and particularly about the regulation of the LTR. I used genetic, molecular, and with the help of collaborators, evolutionary and developmental approaches to expand our understanding of the PMR not only by identifying new members of the LTR pathway but also by identifying conserved families of reproductive proteins that Sfps may act through to regulate behavioral change.

In Chapter 2, I present the results from Fertility/Fecundity Assays aimed at determining whether the M13 family of proteases known as Neprilysins has a conserved role in reproduction in *Drosophila melanogaster* and might serve as targets of Sfp regulation in the female. This work was done in collaboration with Patrick Callaerts and Carmen Francis at the University of Leuven in Belgium. I selected 5 Neprilysin genes (*Nep1, Nep2, Nep3, Nep4, and Nep5)* for their similarity to canonical *Neps* in mammals and their expression in the reproductive tract and the central nervous system. I knocked down each gene in either the male or the female using RNAi and tissue specific drivers where appropriate. Using Fertility/Fecundity Assays I looked for

changes in female egg-laying or hatchability (#progeny/#eggs) in mates of Nep RNAi males or RNAi females compared to controls. Using these methods I was able to determine that knockdown of *Nep1* in both males and females resulted in reduced egg-laying, analogous to the reduction in litter size observed for Neprilysins in the mouse model [72,110,111]. In addition I determined that the insect-specific *Nep2* plays a substantial role in the female, regulating long term egg-laying and hatchability. In addition, I used fluorescence microscopy and immunohistochemistry to determine that the hatchability defect for Nep2 null females is due to early embryonic arrest. Further, this chapter includes sperm count and sperm competition data that support a role for Nep2 in female based regulation of sperm utilization. Combined, my results on Nep1 and Nep2 suggest that *Drosophila* Neps may play an influential role in regulating the PMR, particularly in the female. In addition, the importance of Neps in reproduction, more specifically in total progeny production, of both males and females is conserved in mice and *Drosophila*; establishing the fly as a model for studying the function of *Neps*.

In Chapter 3, I discuss the evolution of a gene family that encodes both female and male specific reproductive proteins in *Drosophila.* This work, done in collaboration with other current and past members of the Wolfner Lab including Laura Sirot, Geoff Findlay, Dorina Frasheri, and Frank Avila, is aimed at evaluating one possible source of Sfp genes, the duplication and cooption of a female gene necessary for the PMR.. The gene family contains the putative serine-proteases CG9897 and CG32834, which are female expressed, and CG32833 which is expressed in the male. We demonstrated that the ancestral copy of this gene family was likely female specific. I present results from fertility fecundity assays using RNAi that support a role for the female expressed gene CG32834 in regulating egg laying in the STR while both female genes,

CG32834 and CG9897, contribute to the regulation of long term female receptivity. Further, the male expressed duplicate CG32833 is also essential for normal egg-laying in females but does not function in regulating female receptivity. Our finding that a seminal fluid protein, with detectable influence on the PMR, arose from the duplication of a female-specific reproductive protein adds to a growing body of work on the evolutionary origin of Sfps.

Chapter 4 centers around the use of ERC to identify additional members of the LTR pathway in both male and female *Drosophila*. This work was done in collaboration with Geoff Findlay and Nathan Clark who used ERC to identify candidate genes in the male and the female. Using RNAi to knock down each candidate we were able to identify a subset of these genes which are important for the storage of SP and the maintenance of the LTR. Further, using western blots to look at the transfer, stability, and persistence of known LTR members we were able to place all but one of these genes in the pathway. My major contribution to this chapter is the work on *intrepid* (CG12558), a serine protease that is necessary for long term storage of SP. *Intrepid* was originally identified through a screen of Sfps that will be discussed in Appendix C. In total, using the ERC method we were able to expand the number of genes known to be important for the LTR, adding three new male and two new female genes to the network.

In Chapter 5 I looked at the contribution of the secondary cells of the male accessory gland to the regulation of the PMR. This work was done in collaboration with François Karch, Dragan Gligorov, and Robert Maeda at the University of Geneva. The Karch Lab demonstrated that the Hox gene *Abd-B* is expressed in the secondary cells of the male accessory gland. They showed that this expression depends on the *iab-6* regulatory region of *Abd-B* and that deletion of a 2.1kb section of *iab-6,* near the *fab-7* boundary, results in loss of *Abd-B* expression in the secondary cells. Further, this region is sufficient to drive expression of GFP in the secondary

cells. Using the *iab-6* deletion allele, *iab-6*^{cocu}, I determined that the secondary cells are necessary for maintenance of the LTR but that the initiation of the STR is not affected in our mutant. The *iab-6*^{cocu} mutant negatively impacts egg-laying, receptivity suppression, and SP storage as well as influencing sperm storage and sperm competition outcomes. Further, several known LTR pathway members are abnormal in molecular weight (CG1656, CG1652, and CG17575), stability (CG9997), or abundance (CG17575) in *iab-6*^{cocu} males. In addition, a driver produced by the region identified in our mutant allele allowed me to knock down gene expression specifically in the secondary cells. I determined that three known LTR pathway genes, CG17575, CG1656, and CG1652 are expressed in this cell type whereas Seminase and CG9997 are main cell specific.

In Chapter 6 I look deeper into the gene expression differences between the *iab-6*^{cocu} mutant and controls to identify individual genes that contribute to the phenotypes observed in our mutant. I selected candidates based on RNA-seq data obtained by our collaborators, Dragan Gligorov and François Karch and I knocked down these candidate genes in the secondary cells using the *iab-6D1*-GAL4 driver discussed in the previous chapter. Of the 19 genes tested I identified 8 genes whose expression in the secondary cells is necessary for normal egg-laying in the LTR, 7 of which also play a role in regulating female receptivity. Further, two of these genes are necessary for the proper glycosylation of CG1656 and CG1652 and three of these genes are needed to maintain normal secondary cell morphology. This work is ongoing and as yet we have been unable to tie CG9997 instability or CG17575 abundance to the LTR related effects seen in mates of *iab-6*^{cocu} males. Surprisingly, only one of our 8 new LTR associated proteins are known to be transferred to females, suggesting that the *iab-6*^{cocu} mutant may not primarily affect Sfps

directly but instead might work through disrupting other cellular functions such as vacuole associated secretion.

Appendix A contains a full excerpt of the text I contributed to the review article mentioned in this introduction[1]: Sirot, L.K., LaFlamme, B.A., Sitnik, J.L., Rubinstein, C.D., Avila, F.W., Chow, C.Y., and Wolfner, M.F. (2009). Molecular social interactions: *Drosophila melanogaster* seminal fluid proteins as a case study. In Advances in Genetics , M.B. Sokolowski, ed. (San Diego: Elsevier Academic Press Inc), pp. 23–56.

Appendix B contains a full list that I have compiled, of all known Sfps in *Drosophila melanogaster* to date, including information about whether they are known to be transferred to the female.

Appendix C reports results from a screen that I began during my rotation. In it, I knocked down candidate genes based on their expression in the accessory gland and tested them for roles in regulating the female PMR.

REFERENCES

1. Sirot LK, LaFlamme BA, Sitnik JL, Rubinstein CD, Avila FW, et al. (2009) Molecular social interactions: Drosophila melanogaster seminal fluid proteins as a case study. Adv Genet 68: 23-56.

2. Ratto MH, Leduc YA, Valderrama XP, van Straaten KE, Delbaere LT, et al. (2012) The nerve of ovulation-inducing factor in semen. Proc Natl Acad Sci U S A 109: 15042-15047.

3. Adams GP, Ratto MH (2013) Ovulation-inducing factor in seminal plasma: a review. Anim Reprod Sci 136: 148-156.

4. den Boer SP, Boomsma JJ, Baer B (2009) Honey bee males and queens use glandular secretions to enhance sperm viability before and after storage. J Insect Physiol 55: 538-543.

5. Collins AM, Caperna TJ, Williams V, Garrett WM, Evans JD (2006) Proteomic analyses of male contributions to honey bee sperm storage and mating. Insect Mol Biol 15: 541-549.

6. Chapman T, Bangham J, Vinti G, Seifried B, Lung O, et al. (2003) The sex peptide of Drosophila melanogaster: female post-mating responses analyzed by using RNA interference. Proc Natl Acad Sci U S A 100: 9923-9928.

7. Liu H, Kubli E (2003) Sex-peptide is the molecular basis of the sperm effect in Drosophila melanogaster. Proc Natl Acad Sci U S A 100: 9929-9933.

8. Chapman T, Herndon LA, Heifetz Y, Partridge L, Wolfner MF (2001) The Acp26Aa seminal fluid protein is a modulator of early egg hatchability in Drosophila melanogaster. Proc Biol Sci 268: 1647-1654.

9. Adams EM, Wolfner MF (2007) Seminal proteins but not sperm induce morphological changes in the Drosophila melanogaster female reproductive tract during sperm storage. J Insect Physiol 53: 319-331.

10. Avila FW, Wolfner MF (2009) Acp36DE is required for uterine conformational changes in mated Drosophila females. Proc Natl Acad Sci U S A 106: 15796-15800.

11. Lawniczak MK, Begun DJ (2004) A genome-wide analysis of courting and mating responses in Drosophila melanogaster females. Genome 47: 900-910.

12. McGraw LA, Gibson G, Clark AG, Wolfner MF (2004) Genes regulated by mating, sperm, or seminal proteins in mated female Drosophila melanogaster. Curr Biol 14: 1509-1514.

13. Peng J, Zipperlen P, Kubli E (2005) Drosophila sex-peptide stimulates female innate immune system after mating via the Toll and Imd pathways. Curr Biol 15: 1690-1694.

14. Kapelnikov A, Zelinger E, Gottlieb Y, Rhrissorrakrai K, Gunsalus KC, et al. (2008) Mating induces an immune response and developmental switch in the Drosophila oviduct. Proc Natl Acad Sci U S A 105: 13912-13917.

15. Boswell RE, Mahowald AP (1985) tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. Cell 43: 97-104.

16. Kalb JM, DiBenedetto AJ, Wolfner MF (1993) Probing the function of Drosophila melanogaster accessory glands by directed cell ablation. Proc Natl Acad Sci U S A 90: 8093-8097.

17. Chapman T, Liddle LF, Kalb JM, Wolfner MF, Partridge L (1995) Cost of mating in Drosophila melanogaster females is mediated by male accessory gland products. Nature 373: 241-244.

18. Xue L, Noll M (2000) Drosophila female sexual behavior induced by sterile males showing copulation complementation. Proc Natl Acad Sci U S A 97: 3272-3275.

19. Xue L, Noll M (2002) Dual role of the Pax gene paired in accessory gland development of Drosophila. Development 129: 339-346.

20. Lung O, Wolfner MF (2001) Identification and characterization of the major Drosophila melanogaster mating plug protein. Insect Biochem Mol Biol 31: 543-551.

21. Bretman A, Lawniczak MK, Boone J, Chapman T (2010) A mating plug protein reduces early female remating in Drosophila melanogaster. J Insect Physiol 56: 107-113.

22. Alonso-Pimentel H, Tolbert LP, Heed WB (1994) Ultrastructural examination of the insemination reaction in Drosophila. Cell Tissue Res 275: 467-479.

23. Moshitzky P, Fleischmann I, Chaimov N, Saudan P, Klauser S, et al. (1996) Sex-peptide activates juvenile hormone biosynthesis in the Drosophila melanogaster corpus allatum. Arch Insect Biochem Physiol 32: 363-374.

24. Heifetz Y, Lung O, Frongillo EA, Jr., Wolfner MF (2000) The Drosophila seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. Curr Biol 10: 99-102.

25. Herndon LA, Wolfner MF (1995) A Drosophila seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating. Proc Natl Acad Sci U S A 92: 10114-10118.

26. Park M, Wolfner MF (1995) Male and female cooperate in the prohormone-like processing of a Drosophila melanogaster seminal fluid protein. Dev Biol 171: 694-702.

27. Lee HG, Rohila S, Han KA (2009) The octopamine receptor OAMB mediates ovulation via Ca2+/calmodulin-dependent protein kinase II in the Drosophila oviduct epithelium. PLoS One 4: e4716.

28. Peng J, Chen S, Busser S, Liu H, Honegger T, et al. (2005) Gradual release of sperm bound sex-peptide controls female postmating behavior in Drosophila. Curr Biol 15: 207-213.

29. Avila FW, Ravi Ram K, Bloch Qazi MC, Wolfner MF (2010) Sex peptide is required for the efficient release of stored sperm in mated Drosophila females. Genetics 186: 595-600.

30. Manning A (1967) The control of sexual receptivity in Drosophila. Animal Behavior: 239-250.

31. Manning A (1962) A sperm factor affecting the receptivity of Drosophila melanogaster females. . Nature 194: 252-253.

32. Domanitskaya EV, Liu H, Chen S, Kubli E (2007) The hydroxyproline motif of male sex peptide elicits the innate immune response in Drosophila females. FEBS J 274: 5659-5668.

33. Yapici N, Kim YJ, Ribeiro C, Dickson BJ (2008) A receptor that mediates the post-mating switch in Drosophila reproductive behaviour. Nature 451: 33-37.

34. Aigaki T, Fleischmann I, Chen PS, Kubli E (1991) Ectopic expression of sex peptide alters reproductive behavior of female D. melanogaster. Neuron 7: 557-563.

35. Ram KR, Wolfner MF (2009) A network of interactions among seminal proteins underlies the long-term postmating response in Drosophila. Proc Natl Acad Sci U S A 106: 15384-15389.

36. Ram KR, Wolfner MF (2007) Sustained post-mating response in Drosophila melanogaster requires multiple seminal fluid proteins. PLoS Genet 3: e238.

37. LaFlamme BA, Ram KR, Wolfner MF (2012) The Drosophila melanogaster seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. PLoS Genet 8: e1002435.

38. Kubli E (2003) Sex-peptides: seminal peptides of the Drosophila male. Cell Mol Life Sci 60: 1689-1704.

39. Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39: 715-720.

40. Rezaval C, Pavlou HJ, Dornan AJ, Chan YB, Kravitz EA, et al. (2012) Neural circuitry underlying Drosophila female postmating behavioral responses. Curr Biol 22: 1155-1165.

41. Salvemini M, Polito C, Saccone G (2010) Fruitless alternative splicing and sex behaviour in insects: an ancient and unforgettable love story? J Genet 89: 287-299.

42. Yang CH, Rumpf S, Xiang Y, Gordon MD, Song W, et al. (2009) Control of the postmating behavioral switch in Drosophila females by internal sensory neurons. Neuron 61: 519-526.

43. Siwicki KK, Kravitz EA (2009) Fruitless, doublesex and the genetics of social behavior in Drosophila melanogaster. Curr Opin Neurobiol 19: 200-206.

44. Swanson WJ, Wong A, Wolfner MF, Aquadro CF (2004) Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. Genetics 168: 1457-1465.

45. Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proc Natl Acad Sci U S A 98: 7375-7379.

46. Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. Nat Rev Genet 3: 137-144.

47. Findlay GD, Yi X, Maccoss MJ, Swanson WJ (2008) Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. PLoS Biol 6: e178.

48. Findlay GD, MacCoss MJ, Swanson WJ (2009) Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. Genome Res 19: 886-896.

49. Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, et al. (2007) Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. Genetics 177: 1321-1335.

50. Wong A, Turchin MC, Wolfner MF, Aquadro CF (2008) Evidence for positive selection on Drosophila melanogaster seminal fluid protease homologs. Mol Biol Evol 25: 497-506.

51. Mueller JL, Ripoll DR, Aquadro CF, Wolfner MF (2004) Comparative structural modeling and inference of conserved protein classes in Drosophila seminal fluid. Proc Natl Acad Sci U S A 101: 13542-13547.

52. Laflamme BA, Wolfner MF (2013) Identification and function of proteolysis regulators in seminal fluid. Mol Reprod Dev 80: 80-101.

53. Monsma SA, Harada HA, Wolfner MF (1990) Synthesis of two Drosophila male accessory gland proteins and their fate after transfer to the female during mating. Dev Biol 142: 465-475.

54. Steele JG, Dalton BA, Hoffman H, Underwood PA, Rathjen D (1985) Monoclonal antibodies to nerve growth factor from bovine seminal plasma. Mol Immunol 22: 1061-1072.

55. Heinrich G, Meyer TE (1988) Nerve growth factor (NGF) is present in human placenta and semen, but undetectable in normal and Paget's disease blood: measurements with an anti-mouse-NGF enzyme immunoassay using a recombinant human NGF reference. Biochem Biophys Res Commun 155: 482-486.

56. Bogle OA, Ratto MH, Adams GP (2012) Ovulation-inducing factor (OIF) induces LH secretion from pituitary cells. Anim Reprod Sci 133: 117-122.

57. Turner AJ, Isaac RE, Coates D (2001) The neprilysin (NEP) family of zinc metalloendopeptidases: genomics and function. Bioessays 23: 261-269.

58. Turner AJ, Brown CD, Carson JA, Barnes K (2000) The neprilysin family in health and disease. Adv Exp Med Biol 477: 229-240.

59. Bland ND, Pinney JW, Thomas JE, Turner AJ, Isaac RE (2008) Bioinformatic analysis of the neprilysin (M13) family of peptidases reveals complex evolutionary and functional relationships. BMC Evol Biol 8: 16.

60. Segura J, Ruilope LM (2011) Dual-acting angiotensin receptor-neprilysin inhibition. Curr Hypertens Rep 13: 74-78.

61. Wick MJ, Buesing EJ, Wehling CA, Loomis ZL, Cool CD, et al. (2011) Decreased neprilysin and pulmonary vascular remodeling in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 183: 330-340.

62. Klein C, Patte-Mensah C, Taleb O, Bourguignon JJ, Schmitt M, et al. (2013) The neuroprotector kynurenic acid increases neuronal cell survival through neprilysin induction. Neuropharmacology.

63. Mulder SD, Veerhuis R, Blankenstein MA, Nielsen HM (2012) The effect of amyloid associated proteins on the expression of genes involved in amyloid-beta clearance by adult human astrocytes. Exp Neurol 233: 373-379.

64. Wong SS, Sun NN, Fastje CD, Witten ML, Lantz RC, et al. (2011) Role of neprilysin in airway inflammation induced by diesel exhaust emissions. Res Rep Health Eff Inst: 3-40.

65. Maguer-Satta V, Besancon R, Bachelard-Cascales E (2011) Concise review: neutral endopeptidase (CD10): a multifaceted environment actor in stem cells, physiological mechanisms, and cancer. Stem Cells 29: 389-396.

66. Smollich M, Gotte M, Yip GW, Yong ES, Kersting C, et al. (2007) On the role of endothelin-converting enzyme-1 (ECE-1) and neprilysin in human breast cancer. Breast Cancer Res Treat 106: 361-369.

67. Pennefather JN, Patak E, Pinto FM, Candenas ML (2004) Mammalian tachykinins and uterine smooth muscle: the challenge escalates. Eur J Pharmacol 500: 15-26.

68. Pennefather JN, Lecci A, Candenas ML, Patak E, Pinto FM, et al. (2004) Tachykinins and tachykinin receptors: a growing family. Life Sci 74: 1445-1463.

69. Patak E, Pennefather JN, Fleming A, Story ME (2002) Functional characterization of tachykinin NK1 receptors in the mouse uterus. Br J Pharmacol 137: 1247-1254.

70. Patak EN, Pennefather JN, Story ME (2000) Effects of tachykinins on uterine smooth muscle. Clin Exp Pharmacol Physiol 27: 922-927.

71. Patak EN, Ziccone S, Story ME, Fleming AJ, Lilley A, et al. (2000) Activation of neurokinin NK(2) receptors by tachykinin peptides causes contraction of uterus in pregnant women near term. Mol Hum Reprod 6: 549-554.

72. Pintado CO, Pinto FM, Pennefather JN, Hidalgo A, Baamonde A, et al. (2003) A role for tachykinins in female mouse and rat reproductive function. Biol Reprod 69: 940-946.

73. Pinto FM, Armesto CP, Magraner J, Trujillo M, Martin JD, et al. (1999) Tachykinin receptor and neutral endopeptidase gene expression in the rat uterus: characterization and regulation in response to ovarian steroid treatment. Endocrinology 140: 2526-2532.

74. Baer B, Zareie R, Paynter E, Poland V, Millar AH (2012) Seminal fluid proteins differ in abundance between genetic lineages of honeybees. J Proteomics 75: 5646-5653.

75. Baer B, Heazlewood JL, Taylor NL, Eubel H, Millar AH (2009) The seminal fluid proteome of the honeybee Apis mellifera. Proteomics 9: 2085-2097.

76. Skidgel RA, Erdos EG (2004) Angiotensin converting enzyme (ACE) and neprilysin hydrolyze neuropeptides: a brief history, the beginning and follow-ups to early studies. Peptides 25: 521-525.

77. Rylett CM, Walker MJ, Howell GJ, Shirras AD, Isaac RE (2007) Male accessory glands of Drosophila melanogaster make a secreted angiotensin I-converting enzyme (ANCE), suggesting a role for the peptide-processing enzyme in seminal fluid. J Exp Biol 210: 3601-3606.

78. Hurst D, Rylett CM, Isaac RE, Shirras AD (2003) The drosophila angiotensin-converting enzyme homologue Ance is required for spermiogenesis. Dev Biol 254: 238-247.

79. Coates D, Siviter R, Isaac RE (2000) Exploring the Caenorhabditis elegans and Drosophila melanogaster genomes to understand neuropeptide and peptidase function. Biochem Soc Trans 28: 464-469.

80. Poels J, Birse RT, Nachman RJ, Fichna J, Janecka A, et al. (2009) Characterization and distribution of NKD, a receptor for Drosophila tachykinin-related peptide 6. Peptides 30: 545-556.

81. Yamamoto MT, Takemori N (2010) Proteome profiling reveals tissue-specific protein expression in the male reproductive system of Drosophila melanogaster. Fly (Austin) 4: 36-39.

82. Takemori N, Yamamoto MT (2009) Proteome mapping of the Drosophila melanogaster male reproductive system. Proteomics 9: 2484-2493.

83. Wolfner MF, Harada HA, Bertram MJ, Stelick TJ, Kraus KW, et al. (1997) New genes for male accessory gland proteins in Drosophila melanogaster. Insect Biochem Mol Biol 27: 825-834.

84. Lee HG, Seong CS, Kim YC, Davis RL, Han KA (2003) Octopamine receptor OAMB is required for ovulation in Drosophila melanogaster. Dev Biol 264: 179-190.

85. Wong A, Albright SN, Giebel JD, Ram KR, Shuqing J, et al. (2008) A Role for Acp29AB, a Predicted Seminal Fluid Lectin, in Female Sperm Storage in Drosophila melanogaster. Genetics 180: 11p.

86. Lung O, Tram U, Finnerty CM, Eipper-Mains MA, Kalb JM, et al. (2002) The Drosophila melanogaster seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. Genetics 160: 211-224.

87. Ravi Ram K, Ji S, Wolfner MF (2005) Fates and targets of male accessory gland proteins in mated female Drosophila melanogaster. Insect Biochem Mol Biol 35: 1059-1071.

88. Chen PS, Buhler R (1970) [Isolation and function of sex peptide in Drosophila melanogaster]. Rev Suisse Zool 77: 548-554.

89. Chen PS, Buhler R (1970) Paragonial substance (sex peptide) and other free ninhydrin-positive components in male and female adults of Drosophila melanogaster. J Insect Physiol 16: 615-627.

90. Chen PS, Stumm-Zollinger E, Aigaki T, Balmer J, Bienz M, et al. (1988) A male accessory gland peptide that regulates reproductive behavior of female D. melanogaster. Cell 54: 291-298.

91. Baumann H, Wilson KJ, Chen PS, Humbel RE (1975) The amino-acid sequence of a peptide (PS-1) from Drosophila funebris: a paragonial peptide from males which reduces the receptivity of the female. Eur J Biochem 52: 521-529.

92. Schnakenberg SL, Matias WR, Siegal ML (2011) Sperm-storage defects and live birth in Drosophila females lacking spermathecal secretory cells. PLoS Biol 9: e1001192.

93. Anderson RC (1945) A Study of the Factors Affecting Fertility of Lozenge Females of Drosophila Melanogaster. Genetics 30: 280-296.

94. Allen AK, Spradling AC (2008) The Sf1-related nuclear hormone receptor Hr39 regulates Drosophila female reproductive tract development and function. Development 135: 311-321.

95. Sun J, Spradling AC (2013) Ovulation in Drosophila is controlled by secretory cells of the female reproductive tract. Elife 2: e00415.

96. Clark NL, Alani E, Aquadro CF (2012) Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res 22: 714-720.

97. Tabach Y, Billi AC, Hayes GD, Newman MA, Zuk O, et al. (2013) Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. Nature 493: 694-698.

98. Lung O, Kuo L, Wolfner MF (2001) Drosophila males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. J Insect Physiol 47: 617-622.

99. Bairati A (1968) Structure and ultrastructure of the male reproductive system in *Drosophila melanogaster*. Monitore zoologico italiano 2: 105-182.

100. Bertram MJ, Akerkar GA, Ard RL, Gonzalez C, Wolfner MF (1992) Cell type-specific gene expression in the Drosophila melanogaster male accessory gland. Mech Dev 38: 33-40.

101. DiBenedetto AJ, Harada HA, Wolfner MF (1990) Structure, cell-specific expression, and mating-induced regulation of a Drosophila melanogaster male accessory gland gene. Dev Biol 139: 134-148.

102. Brissova M, Fowler MJ, Nicholson WE, Chu A, Hirshberg B, et al. (2005) Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. J Histochem Cytochem 53: 1087-1097.

103. Cornwall GA (2009) New insights into epididymal biology and function. Hum Reprod Update 15: 213-227.

104. Cheung KH, Leung GP, Leung MC, Shum WW, Zhou WL, et al. (2005) Cell-cell interaction underlies formation of fluid in the male reproductive tract of the rat. J Gen Physiol 125: 443-454.

105. Pietrement C, Sun-Wada GH, Silva ND, McKee M, Marshansky V, et al. (2006) Distinct expression patterns of different subunit isoforms of the V-ATPase in the rat epididymis. Biol Reprod 74: 185-194.

106. Kujala M, Hihnala S, Tienari J, Kaunisto K, Hastbacka J, et al. (2007) Expression of ion transport-associated proteins in human efferent and epididymal ducts. Reproduction 133: 775-784.

107. Hermo L (1995) Structural features and functions of principal cells of the intermediate zone of the epididymis of adult rats. Anat Rec 242: 515-530.

108. Minami R, Wakabayashi M, Sugimori S, Taniguchi K, Kokuryo A, et al. (2012) The homeodomain protein defective proventriculus is essential for male accessory gland development to enhance fecundity in Drosophila. PLoS One 7: e32302.

109. Gligorov D, Sitnik JL, Maeda RK, Wolfner MF, Karch F (2013) A novel function for the Hox gene Abd-B in the male accessory gland regulates the long-term female post-mating response in Drosophila. PLoS Genet 9: e1003395.

110. Carpentier M, Guillemette C, Bailey JL, Boileau G, Jeannotte L, et al. (2004) Reduced fertility in male mice deficient in the zinc metallopeptidase NL1. Mol Cell Biol 24: 4428-4437.

111. Ghaddar G, Ruchon AF, Carpentier M, Marcinkiewicz M, Seidah NG, et al. (2000) Molecular cloning and biochemical characterization of a new mouse testis soluble-zinc-metallopeptidase of the neprilysin family. Biochem J 347: 419-429.

CHAPTER 2

DROSOPHILA ORTHOLOGS OF MAMMALIAN NEPRILYSIN FAMILY PROTEINS PLAY IMPORTANT ROLES IN REPRODUCTION[2]

1.1 INTRODUCTION

Proteases play key roles in diverse physiological systems. One such family of metalloproteases, the M13 class of neutral endopeptidases, consists mainly of membrane bound zinc proteases that are involved in the processing of neuropeptides and peptide hormones [reviewed in 1,2,3]. In mammals, seven members of this family have been identified, of which neprilysin (NEP) and endothelin converting enzyme (ECE) are the best-studied. These proteins have been implicated in various diseases including cardiovascular disease [4,5], Alzheimer's disease [6,7], inflammation and inflammatory disorders [8], and cancer [9,10]. In addition to their role in disease, NEPs are essential for development and reproduction in mammals, but how they affect these processes is poorly understood. *Neprilysin-2*, called *NL1* in mice, is highly expressed in the testis. *NL1*-deficient males sire fewer pups, even though spermatogenesis appears to be unaffected [11]. In female rats and mice, degradation of tachykinins by NEP in the uterus is essential for controlling uterine contractions, and an inability to degrade tachykinins is associated with a reduction in litter size [12,13].

To understand the physiological roles of neprilysins, we focused on this gene family in the genetically tractable model *Drosophila melanogaster*. The *D. melanogaster* genome has 24 NEP-like genes, most of which are actively transcribed [3,14,15]. However, little is

[2] A version of this chapter was submitted to Genetics as "Sitnik, J., Francis, C., Hens, K., Huybrechts, R., Wolfner, M., and Callaerts, P., "*Drosophila* orthologs of mammalian Neprilysin family proteins play important roles in reproduction." Figures and data contributed by Carmen Francis and Patrick Callaerts are denoted in the figure legends. This is primarily the *in situ* hybridization results and some of the alignment work. A special thanks to Geoff Findlay for helping with Figure 2.1.

known about their roles *in vivo*. Neprilysin-like activity has been detected in extracts of larval imaginal discs and of neuronal membranes from larval and adult heads of *Drosophila* [16,17]. At least two *Drosophila* genes, *Nep2* [18] and *Nep4* [19] are active proteases with specific substrate affinities that can be inhibited with the M13-specific peptidase inhibitors thiorphan and phosphoramidon. Nep2 has been shown to cleave locustatachykinin-1 (LomTK-1) and *Drosophila* tachykinins *in vitro* at a Gly-Val peptide bond [20]. Roles for *Drosophila Nep2* in renal function and reproduction have been suggested based on its expression in malpighian tubules and the reproductive organs of both sexes [14,20].

Here, we examined the phylogeny of *Drosophila* Neprilysin proteins and analyzed the function and the expression patterns of a sub family, containing *Neprilysin1 (Nep1), Neprilysin2 (Nep2), Neprilysin3 (Nep3), Neprilysin4 (Nep4)* and *Neprilysin5 (Nep5*), whose expression pattern is most similar to the canonical mammalian Neprilysin. Our mutational and RNAi studies revealed that *Drosophila* Neprilysins are important in egg-production and also for regulating sperm use in mated females.

2.2 RESULTS

Sequence and phylogenetic analysis of Neprilysins

Twenty four peptidase sequences encoded in the *Drosophila melanogaster* genome are classified as M13 metallopeptidases based on gene prediction, sequence homology and searches for known active site regions using the MEROPS database [21]. We created a tree of all 24 M13 class proteins by comparing their protein sequence similarity (see methods). The M13 class proteins fall into three related groups. We were specifically interested in clusters of M13 genes that are expressed in the reproductive tract (RT) of either sex or in the central nervous system (CNS), as either expression pattern may suggest a role in reproduction. To determine which genes fit our criteria we mapped the known expression patterns [Fly

Atlas; 14] for either the RT (as indicated by genome-wide microarray data determined in females for the ovaries and the spermathecae, and in males for the testes and the accessory glands) or the CNS onto the gene tree (Figure 2.1). All but two of the 24 genes (CG9507 and CG4580) have some expression in the RT of either sex. Most genes (19/24) show some expression in female reproductive tracts; only 8/24 are detectably expressed in male reproductive tract tissues. Fourteen genes show female RT expression only, 3/24 show male RT expression only, and 5/24 are expressed in both. (Table 2.1). The high frequency of female RT expressed genes in this family suggests that the function of M13 class proteins is likely important in these tissues, but also suggests the possibility of functional redundancy which could complicate genetic analysis. Thus we decided to focus on candidate genes with the somewhat rarer pattern of male-only or unbiased RT expression. Only one of the clades is enriched for expression in the male. This same clade is also enriched for genes expressed in the CNS. CNS and reproductive tract expression in both sexes is characteristic of the canonical mammalian Neprilysin [22,23]. Thus, based on both the expression pattern (CNS and RT) and similarity to the expression patterns of Neps known to be important for reproduction in mammals, we chose to focus on this sub-group.

We characterized the five genes in this clade: *Nep1 (CG5905), Nep2 (CG9761), Nep3 (CG9565), Nep4 (CG5894)* and *Nep5 (CG6265)*. Figure 2.2 shows a schematic representation of neprilysin and a sequence alignment of the different functional motifs of *D. melanogaster* Nep1-Nep5, the ECE homolog of *Locusta migratoria* (LomECE) and *Homo sapiens* ECE-1, ECE-2 and neprilysin.[3]

Expression patterns of Nep1-5

FlyAtlas [14] and RNAseq [24] data suggested that *Neps1-5* are expressed throughout

[3] A full sequence alignment and phylogenetic analysis was done by Patrick Callearts and is not included in this chapter. This information is part of the supplement and a full version will be available when the paper is published.

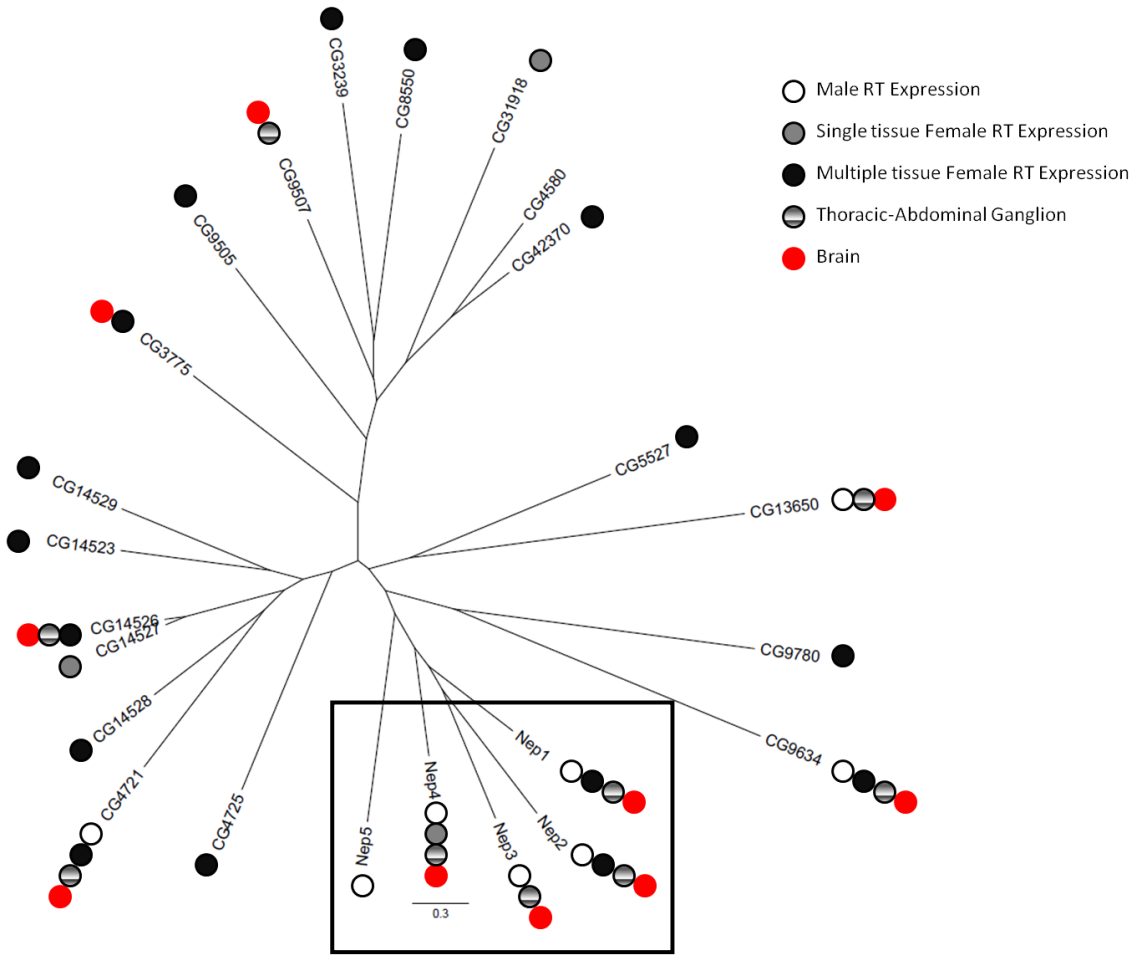


Figure 2.1 Phylogeny and expression of *Drosophila Neprilysins*

A phylogenetic tree of the 24 known *Drosophila melanogaster* M13 class proteases based on protein sequence similarity. The proteins fall into three distinct clades. Mapping the reproductive tract expression of each gene onto the tree reveals broad expression in the female RT (grey and black) and enrichment of male expression (white) in the clade that contains Nep1-5. The same clade also demonstrates enrichment for brain (red) and Abdominal-thoracic ganglion expression (striped).

Table 2.1: Gene expression profiles for *Drosophila Neprilysin* family genes in reproductive organs

Gene	Total Tissues	Virgin Spermatheca	Mated Spermatheca	Ovary	Testis	Male AG	RT Specificity	CNS Expression
CG13650	2	-	-	-	+	+	Male	+
CG14523	2	+	+	-	-	-	Female	-
CG14526	3	+	+	+	-	-	Female	+
CG14527	1	-	+	-	-	-	Female	-
CG14528	2	+	+	-	-	-	Female	-
CG14529	1	+	-	-	-	-	Female	-
CG31918	1	-	-	+	-	-	Female	-
CG3239	2	+	+	-	-	-	Female	-
CG3775	2	+	+	-	-	-	Female	+
CG42370	2	-	+	-	-	-	Female	-
CG4580	0	-	-	-	-	-	None	-
CG4721	3	+	+	-	-	+	Both	+
CG4725	2	+	+	-	-	-	Female	-
CG5527	2	+	+	-	-	-	Female	-
CG8550	2	+	+	-	-	-	Female	-
CG9505	3	+	+	+	-	-	Female	-
CG9507	0	-	-	-	-	-	None	+
CG9634	4	+	+	+	+	-	Both	+
CG9780	2	+	+	-	-	-	Female	-
Nep1	3	+	+	-	+	-	Both	+
Nep2	4	+	-	+	+	-	Both	+
Nep3	1	-	-	-	+	-	Male	+
Nep4	2	+	-	-	+	-	Both	+
Nep5	1	-	-	-	-	+	Male	-

A summary of the gene expression profiles of Drosophila neprilysins in both male and female reproductive organs. Expression data were obtained from Fly atlas [14], positive expression is denoted (+) and lack of expression is denoted with (-). Total tissues refers to the number of reproductive tract tissues in which the gene is found. Mated verses Virgin spermatheca samples were treated separately, since it is known that mating causes differential gene expression in the female spermatheca [25].

A

B

	NAY/FY	HEXXH	EXXXD	CXXW
Dmel_Nep4	NAYY	HELTH	ENIAD	CSVW
Dmel_Nep3	NAYY	HELTH	ENIAD	CEVW
Lmig_ECE	NAYY	HELTH	ENIAD	CEVW
Hsap_ECE1	NAYY	HELTH	ENIAD	CEVW
Hsap_ECE2	NAYY	HELTH	ENIAD	CEVW
Hsap_Nep	NAFY	HEITH	ENIAD	CRVW
Dmel_Nep1	NAYY	HEITH	ENIAD	CSVW
Dmel_Nep2	NAYY	HEITH	ENIAD	CEVW
Dmel_Nep5	NAYH	HELTH	ENIAD	CRIW

Figure 2.2 Conserved binding motifs in *Drosophila*, Human, and Locust *Neprilysins*

A) Schematic representation of neprilysin. Black = cytoplasmic domain, light grey = transmembrane domain, dark grey = extracellular domain. NAY/FY: important for substrate binding; HExxH: zinc binding domain; ExxxD: zinc binding domain; CxxW: sequence critical for protein folding and maturation of the enzyme. **B)** Alignment of NAY/FY, HExxH, ExxxD and CxxW sequences of *Drosophila melanogaster Nep1-5*, *Locusta migratoria* ECE and *Homo sapiens* ECE1-2. This work was contributed by Patrick Callaerts.

development in a variety of tissues. To gain a more precise understanding of the locations and timing of these genes' expression patterns, we performed *in situ* hybridization to look for the expression of each NEP gene in embryos, larvae, and adult flies.

Embryonic expression of Nep1-5

Two of the genes (*Nep3* and *Nep4*) were expressed before embryonic stage 17 (Figure 2.3). *Nep4* RNA was detected as early as stage 12 in two patches of cells per hemisegment (Figure 3A). These have been reported independently to correspond to muscle founder cells [19]. In stage 13 *Nep4* is expressed in two rows of cells that border the amnioserosa (Figure 2.3B), which we identify as the pericardial cells. These cells flank the aligned cardioblast cells of the dorsal vessel. This staining is visible from stage 13 until stage 16 when dorsal closure is finalized and the cardioblasts of each side fuse. At stage 14 *Nep3* is expressed generally in the central nervous system (Figure 2.3C). This staining is visible until stage 17. *Nep4* expression can also be detected in the brain and ventral nerve cord of stage 14 to stage 17 embryos (Figure 2.3D&K). The staining is localized in cells along the longitudinal connectives and transversal commissures of the ventral nerve cord.

All of these five NEP genes are expressed in stage 17 embryos. *Nep1* is expressed in neurons of the peripheral nervous system on the left and right side of the embryo (Figure 2.3E&F), and in the antenno-maxillary complex; which is part of the peripheral nervous system and located at the anterior side of the embryo, in front of the first thoracic segment. *Nep1* RNA was also detected in the anterior of the pharynx and in cells of the embryonic midgut. *Nep2* is strongly expressed in the tracheal system including in the dorsal trunk and the dorsal branches (Figure 2.3G). In the intestinal tract *Nep2* is expressed in the foregut (Figure 2.3H, &I). *Nep2* expression can also be detected in the hindgut (Figure 2.3H) and epidermis (Figure 2.3I). *Nep3* expression remains in the CNS where it becomes more intense

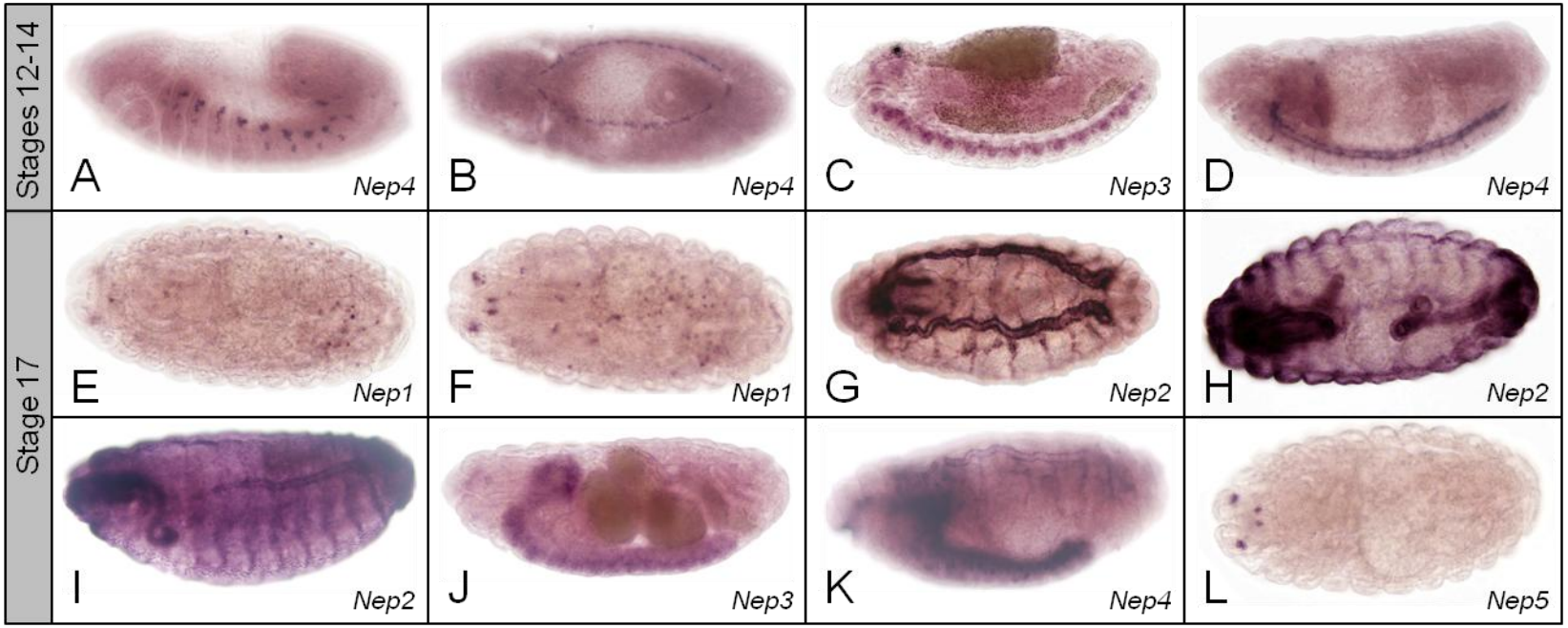


Figure 2.3 Embryonic expression pattern of *neprilysins*

A-B) Embryonic stage 12 Nep4 expression in muscle founder cells (arrows in A) and in pericardial cells (pc). **C)** Embryonic stage 13 Nep3 expression in the ventral nerve cord (vnc). **D)** Embryonic stage 14 Nep4 expression in brain (br) and ventral nerve cord (vnc). **E-F)** Embryonic stage 17 Nep1 expression in peripheral nervous system (arrows), antenno-maxillary complex (circled) and cells in the pharynx (open arrowheads) and midgut (closed arrowheads). **G-I)** Embryonic stage 17 Nep2 expression in dorsal trunk (dt), dorsal branches (db), foregut (fg) and hindgut (hg). **J)** Embryonic stage 17 Nep3 expression in brain (br) and ventral nerve cord (vnc). **K)** Embryonic stage 17 Nep4 expression in brain (br), ventral nerve cord (vnc) and tracheal dorsal trunk (dt). **L)** Embryonic stage 17 Nep5 expression in antenno-maxillary complex (circled) and in the pharynx (open arrowheads). This work was contributed by Carmen Francis and Patrick Callaerts.

in the brain hemispheres (Figure 2.3J). *Nep4* expression is detectable in the dorsal trunk and epidermis (Figure 2.3I). *Nep3* expression remains in the CNS where it becomes more intense in the brain hemispheres (Figure 2.3J). *Nep4* expression is detectable in the dorsal trunk and dorsal branches of the tracheal system and continues to be detected in the brain and ventral nerve cord (Figure 2.3K). Expression of *Nep5* is restricted to four small groups of cells at the anterior of stage 17 embryos (Figure 2.3L).

Larval expression patterns of Nep1-5

Nep1-4 are expressed in the nervous system of third instar larvae (Figure 2.4). *Nep1* is expressed strongly in the mushroom bodies of the brain, neurons in the *pars intercerebralis*, and neurons in the ventral ganglia (Figure 2.4A). *Nep2* is expressed in three neurons of both hemispheres of the larval brain and a limited number of six neurons in the ventral ganglia (Figure 2.4B). Similar to the expression of *Nep3* in embryos; a strong general staining of *Nep3* is detected in the larval brain hemispheres and ventral ganglia. In the hemispheres, the staining is more intense in the central part compared to that in the optic neuropils (Figure 2.4C). The expression of *Nep4* in third instar larvae is restricted to the central nervous system. Based on the size of the cells that are stained in the brain and ventral ganglia we identify the *Nep4*- expressing cells as glia (Figure 2.4D), an observation that was independently made by Meyer *et al.* [19].

Other than the CNS, Neps are expressed in the gut and the malpighian tubules, as well as in developing wing, leg, and eye-antennal discs. More specifically *Nep1* expression is detected in cells of the midgut (Figure 2.4E), wing disc (Figure 2.4I), and leg disc (Figure 2.4J). *Nep2* remains expressed in the foregut, but only in a limited number of cells of the proventriculus (Figure 2.4F). In the eye-antennal disc *Nep2* is expressed anterior to the morphogenetic furrow in the undifferentiated precursor cells of the eye disc and more

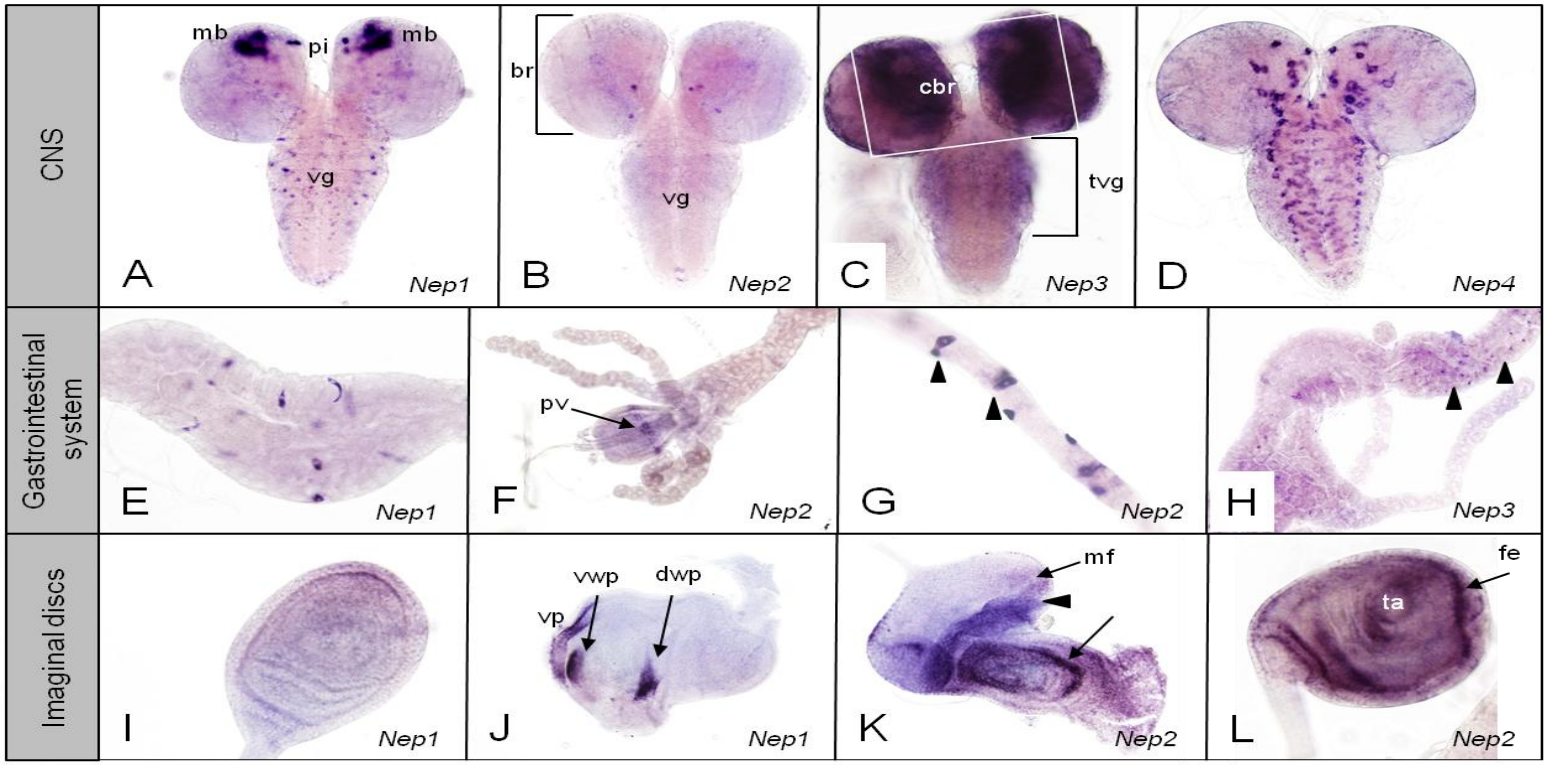


Figure 2.4 Larval expression pattern of neprilysins

A-D) Expression of Nep1-4 in larval CNS. **A)** Nep1 in mushroom bodies (mb), pars intercerebralis (pi) and ventral ganglia (vg). **B)** Nep2 in few cells in the brain (br) and in the ventral ganglia (vg). **C)** Nep in central brain (cbr) and thoracic ventral ganglia (tvg). **D)** Nep4 in glia of the larval CNS. **E-H)** Expression of Nep1-3 in the larval gastrointestinal system. **E)** Nep1 in cells in the midgut. **F-G)** Nep2 in the proventriculus (pv) and in the stellate cells of the Malpighian tubules (arrowheads). **H)** Nep3 in scattered cells in the larval midgut (arrowheads). **I-L)** Expression of Nep1 and Nep2 in larval imaginal discs. **I)** Nep1 in leg disc. **J)** Nep1 in dorsal and ventral wing pouch (dwp - vwp) and ventral pleura (vp) of the wing disc. **K)** Nep 2 in the eye disc anterior (arrowhead) to the morphogenetic furrow (mf) and in the second antennal segment (arrow). **L)** Nep2 in the leg disc femur (fe) and tarsus (ta). This work was contributed by Carmen Francis and Patrick Callaerts.

generally in the antennal part (Figure 2.4K). In the leg discs *Nep2* is expressed in the outer concentric ring, giving rise to the first two segments of the fly leg, and in the central part of the disc (Figure 2.4L). *Nep2* is also expressed in the stellate cells of the larval Malpighian tubules (Figure 2.4G), which perform excretory and osmo-regulatory roles analogous to vertebrate renal tubules [26]. *Nep3* expression is detected in a small number of cells in the larval midgut (Figure 2.4H). We did not detect expression of *Nep5* above background level in third instar larval tissues.

Adult expression patterns for Nep1-5

Consistent with its larval tissue expression pattern, *Nep1* is expressed in the mushroom bodies and neurons of the *pars intercerebralis* of the adult brain (Figure 2.5A) and in cells of the adult midgut (Figure 2.5D). In the male reproductive organs *Nep1* is expressed at the end of the testicular tube near and in the seminal vesicles (Figure 2.5G&H). *Nep2* is detected in neurons of the *pars intercerebralis* and in a limited number of cells in the optic lobes of the brain (Figure 2.5B). In the ventral ganglion a few neurons also show expression of *Nep2* (Figure 2.5C). In the male reproductive organs *Nep2* is expressed in cells at the end of the testicular tube where it meets the seminal vesicle (Figure 2.5I). In the female gonad, strong staining was detected in posterior polar cells and in border cells of stage 8, 9 and 10 follicles (Figure 2.5F). As in larvae, *Nep2* is expressed in the adult Malpighian tubules and more specifically in the stellate cells which are located between the principal cells of the Malpighian tubules (Figure 2.5E). No expression of *Nep3* above background level was detected in adult tissues despite previous reports of broad expression [14,24]. The expression of *Nep4* in adult flies is restricted to the male gonads. Expression of *Nep4* is detected in different parts of the testicular tubes (Figure 2.5J, K, &L). In the apex of the testis the localization of the staining corresponds to the somatic cyst cells that surround the

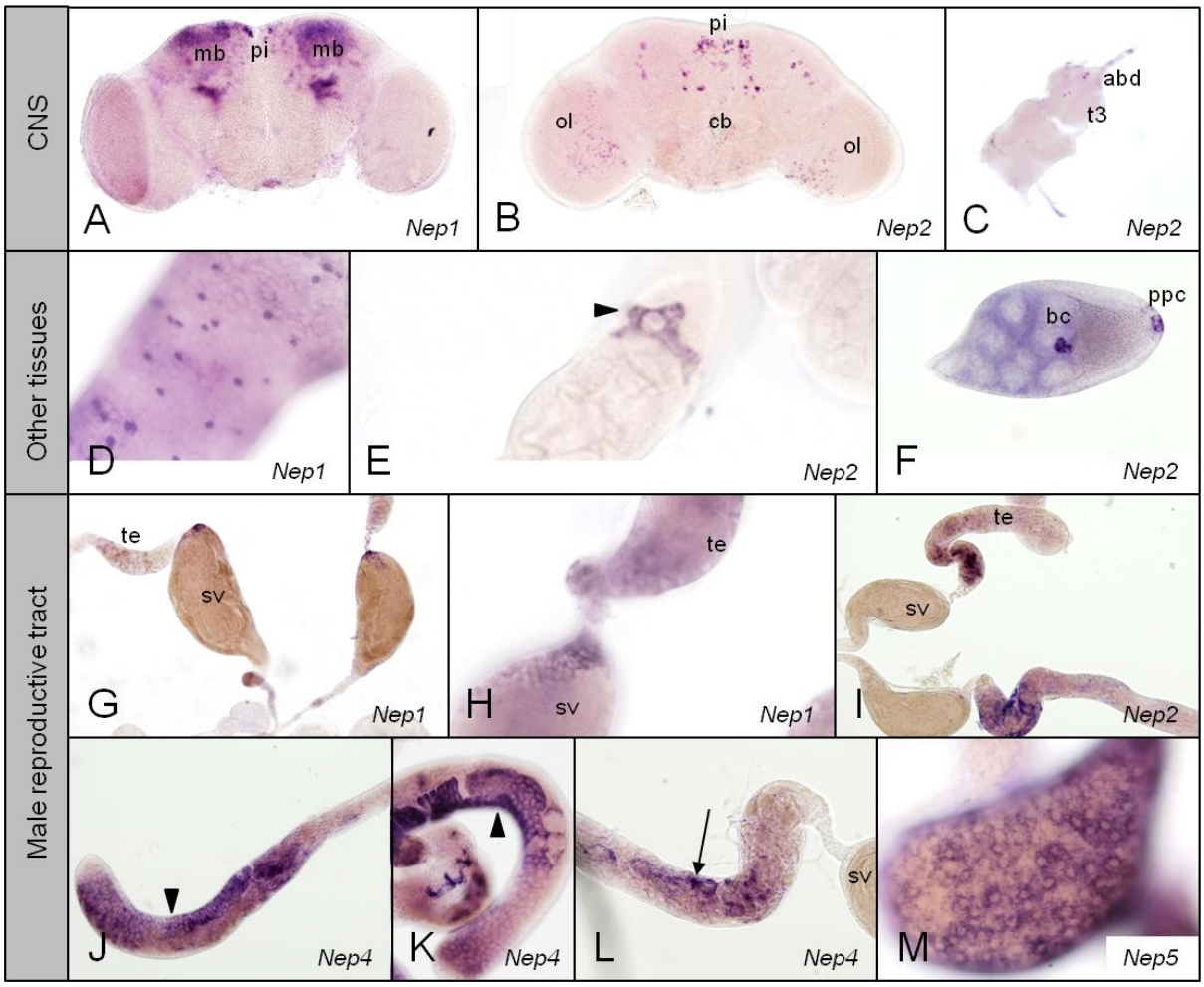


Figure 2.5 Adult expression pattern of *neprilysins*

A-C) Expression of Nep1-2 in the adult CNS. (A) Nep1 in the adult brain mushroom bodies (mb) and pars intercerebralis (pi). **B)** Nep2 in cells in the pars intercerebralis (pi), central brain (cb) and optic lobes (ol). **C)** Nep2 in the third thoracic (t3) and abdominal (abd) neuromere. **D)** Nep1 in adult midgut cells. **E)** Nep2 in adult stellate cells of the Malpighian tubules (arrowhead). (F) Nep2 in border cells (bc) and posterior polar cells (ppc) of a stage 10 ovarian follicle. **G-M)** Expression of neprilysins in the male reproductive tract. **G-H)** Nep1 in the testicular tube (te) and the seminal vesicles (sv). (I) Nep 2 in the part of the testis (te) close to the seminal vesicle (sv). **J-L)** Nep 4 in the somatic cyst cells (arrowheads) and in other cells (arrows) in the part of the testes close to the seminal vesicle (sv). (M) Nep5 in the seminal vesicle. This work was contributed by Carmen Francis and Patrick Callaerts.

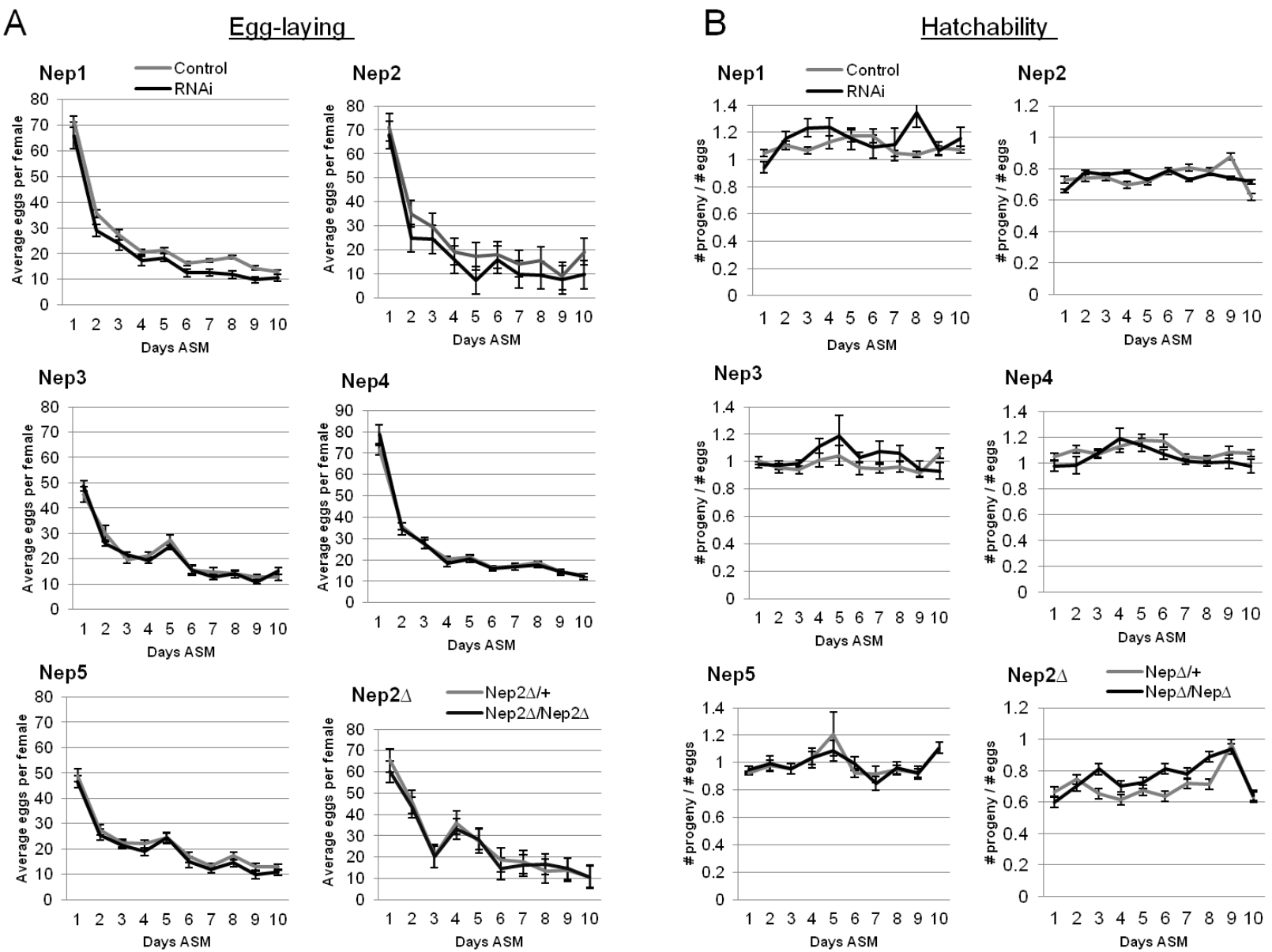
spermatocytes in this part of the testis. *Nep4* is also expressed at the end of the tube close to the contact with the seminal vesicle in cells other than the somatic cyst cells. As is true for *Nep4*, the expression of *Nep5* in adult tissues is also restricted to the male gonads, more specifically in the membrane of the seminal vesicles where mature spermatids are stored after transport from the testicular tubes (Figure 2.5M).

Nep1 plays a role in male fertility.

Combined with the previously reported expression patterns from Fly Atlas and modENCODE, our data showed that in adult flies *Nep4, and Nep5* are expressed predominantly or exclusively in the male reproductive tract. Conversely, *Nep1* and *Nep2* were present at high levels in the reproductive tracts of both sexes as well as in the CNS (Table 2.1). We were able to confirm *Nep3* expression in the male by RT-PCR (data not shown). To test if any of these genes are essential for male fertility we generated knockdown males for *Nep2, Nep3, and Nep5* by crossing the appropriate VDRC line to the ubiquitous driver *tubulin*-GAL4. Knockdown of *Nep1* and *Nep4* was lethal for both sexes using *tubulin*-GAL4, so we relied on *hsp70*-GAL4 (HS-GAL4) instead [27]. Control or knockdown males were mated, in parallel, to virgin females. Egg production and fertility was measured daily over a 10 day period for each female. No differences were seen in the total number of eggs produced by females mated to either *Nep2, Nep3, Nep4,* or *Nep5* knockdown males compared to controls (Figure 2.6A). Mates of *Nep2* knockdown males showed a trend toward reduced fertility. Mates of *Nep1* knockdown males laid significantly fewer eggs than mates of control males (Figure 2.6A). We obtained and tested males from a *Nep2* deletion line to clarify the trend observed in mates of *Nep2* knockdown males. Mates of Nep2 null males showed no difference in egg-laying , suggesting that Nep2 from the male is not essential for this process. While a *Nep1* mutation line exists, it is not a null allele and as such we were unable to verify

Figure 2.6 Egg-laying in mates of *Nep* RNAi males

A) The mean number of eggs laid per female mated to either control males (grey line) or RNAi/null males (black line) over a 10 day period. Only mates of *Nep1* RNAi males laid fewer eggs than mates of control males (*Nep1*: rmANOVA p=0.0047, Control N=20, *Nep1* RNAi N=18). Mates of *Nep2-5* RNAi laid comparable numbers of eggs as control mated females (*Nep2*: rmANOVA p=0.095, Control N=11, *Nep2* RNAi N=14; *Nep3*: rmANOVA p=0.7556 , Control N=17, *Nep3* RNAi N=21; *Nep4*: rmANOVA p=0.9661, Control N=20, *Nep4* RNAi N=17; *Nep5*: rmANOVA p=0.1722, Control N=21, *Nep5* RNAi N=19; Nep2 null: rmANOVA p=0.3448, Control N=18, Nep2 null N=21). **B)** The mean hatchability (#progeny/#eggs) per female for mates of control or RNAi/null males for the egg-laying assays in part A. None of the Neps had a significant effect on hatching rate (*Nep1*: rmANOVA p=0.4751 ; *Nep2*: rmANOVA p=0.4326 ; *Nep3*: rmANOVA p=0.1494; *Nep4*: rmANOVA p=0.1146 ; *Nep5*: rmANOVA p=0.8466; Nep2 null: rmANOVA p=0.3673).


A
Egg-laying
Nep1
Control
RNAi
Average eggs per female
Days ASM
Nep2
Nep3
Nep4
Nep5
Nep2Δ
Nep2Δ/+
Nep2Δ/Nep2Δ
B
Hatchability
#progeny / #eggs
NepΔ/+
NepΔ/NepΔ

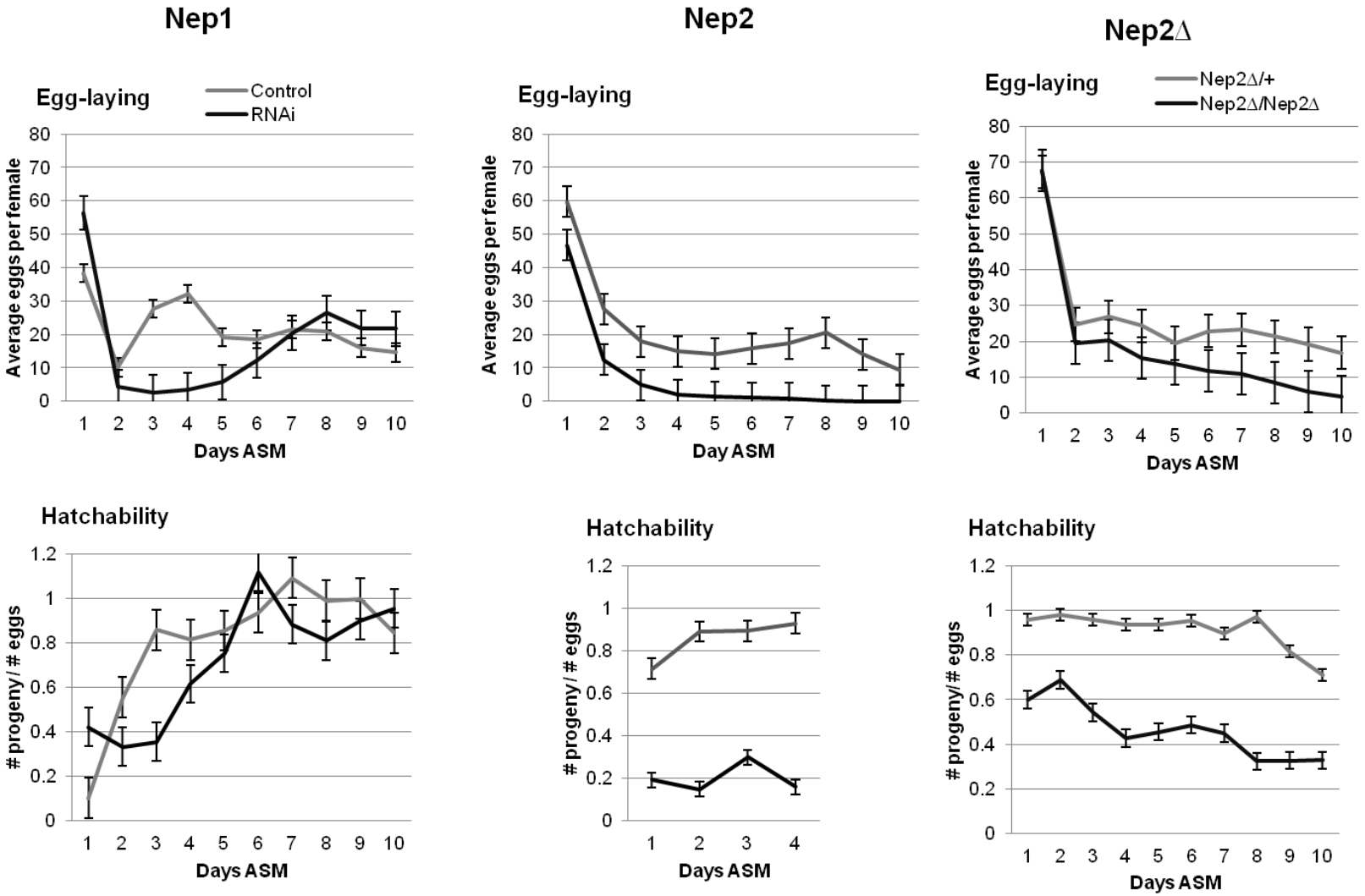
the egg-laying defects seen in mates of *Nep1* knockdown males. The proportion of progeny that eclosed from eggs laid by females (hatchability) mated to any of the five Nep knockdown males was comparable to those of controls (Figure 2.6B). Some small differences in egg-laying were observed for *Nep3* and *Nep5* knockdown males within the first 24 hours after mating, but these differences were not consistently reproducible. This variability may be due to variation in the level of knockdown across males. *Nep2, Nep3, Nep4,* and *Nep5* do not appear to be uniquely essential for male fertility but may perform redundant functions. However, knockdown of *Nep5* was incomplete so we cannot completely rule out a role for this gene at this time. Reduced *Nep1* expression does impact egg-laying in mated females suggesting a role for neprilysins in the reproductive performance of *Drosophila* males.

Nep1 and Nep2 are essential for female fertility

Nep1 and *Nep2* are highly expressed in the female reproductive tract and the CNS based on our expression data (Fig. 2.5) and Fly Atlas [14]. Ubiquitous reduction in the expression of either *Nep1* (using *HS*-GAL4 due to lethality with *tubulin*-GAL4) or *Nep2* (using *tubulin*-GAL4) in females reduced egg-laying compared to control females (Figure 2.7). *Nep1* RNAi females lay significantly more eggs in the first 24 hours after mating and fewer eggs than controls on days 3-5 with an over-all reduction in total eggs laid over the entire 10 day period. Low levels of egg laying in the first 1-2 days after mating is frequently observed in both control and RNAi females following heat shock (Sitnik and Wolfner, unpublished data) and is likely a result of heat stress caused by the heat shock itself. Expression of GAL4 using the HS promoter is transient, resulting in a subsequently transient level of knockdown [28]. While controls recover from the drop in egg-laying associated with exposure to heat shock by 3 days after mating, the Nep1 RNAi females do not recover until

Figure 2.7 Egg-laying in *Nep* RNAi females

A) The mean number of eggs laid by control (grey line) or RNAi/null females (black line) mated to WT males over a 10 day period. Both *Nep1* and *Nep2* RNAi and Nep2 null females lay fewer eggs than controls (*Nep1*: rmANOVA p=0.0435, Control N=11, *Nep1* RNAi N=12; *Nep2*: rmANOVA p=<0.0001*, Control N=12, *Nep2* RNAi N=12 ; Nep2 null: rmANOVA p=<0.0001*, Control N=18, Nep2 null N=15). **B)** The mean hatchability (#progeny/#eggs) per female based on the previous egg-laying assay. The effect of *Nep1* RNAi on hatchability is confounded by heat shock but suggests that there is not a role for *Nep1* in this process (Nep1: rmANOVA p=0.2654). Both *Nep2* RNAi and the Nep2 null females show drastically reduced hatchability (*Nep2*: rmANOVA p=<0.0001*; Nep2 null: rmANOVA p=<0.0001*) suggesting that *Nep2* plays an essential role in this process.


Nep1
Egg-laying
Control
RNAi
Average eggs per female
Days ASM
Hatchability
progeny / # eggs
Days ASM
Nep2
Egg-laying
Average eggs per female
Day ASM
Hatchability
progeny / # eggs
Days ASM
Nep2Δ
Egg-laying
Nep2Δ/+
Nep2Δ/Nep2Δ
Average eggs per female
Days ASM
Hatchability
progeny / # eggs
Days ASM

much later, consistent with efficient knockdown occurring as well as a role for *Nep1* in regulating female egg laying. It is unclear if this reduction in egg-laying indicates a role for *Nep1* in recovery from heat shock directly or on egg-laying specifically. Due to the effects that heat shock has on oogenesis and egg viability in the female [29], it is difficult to determine if *Nep1* plays any role in hatchability.

Females knocked down for *Nep2* laid significantly fewer eggs than controls in total and across all time points (Figure 2.7). In addition, the proportion of eggs that become adult progeny is greatly reduced compared to control females (20% in RNAi females versus 80-90% in controls). It was not possible to calculate hatchability for the entire 10 day period, since egg-laying reached zero for all Nep2 RNAi females by day 5. The reduction in hatchability is unsurprising, as we observed that *Nep2* is present in both the polar and border cells of the follicular epithelium (Fig. 2.5F). Similar results were seen using females that were homozygous for a null mutation of *Nep2* (Figure 2.7). Together these results suggest that *Nep1* and *Nep2* play essential roles in female fertility and fecundity.

Characterization of the hatchability defects in Nep2 null mutants

To test whether the hatchability defect observed for *Nep2* mutant females was due to a failure of the eggs to be fertilized, we examined early embryos (0-1h) laid by these females for the presence of a sperm tail [30]. There was no difference in the percentage of fertilized embryos laid by *Nep2* null females compared to controls (Figure 2.8A). We therefore tested whether *Nep2* is important for early embryogenesis by staining 1.5-3.5h old embryos laid by *Nep2* null females or controls with DAPI and scoring for their stage of embryonic development. While nearly all of the eggs laid by control females contain developing embryos, significantly more of the eggs laid by *Nep2* nulls (close to 50%) only contain a clear polar body rosette (Figure 2.8B). The presence of a polar body rosette is typical for

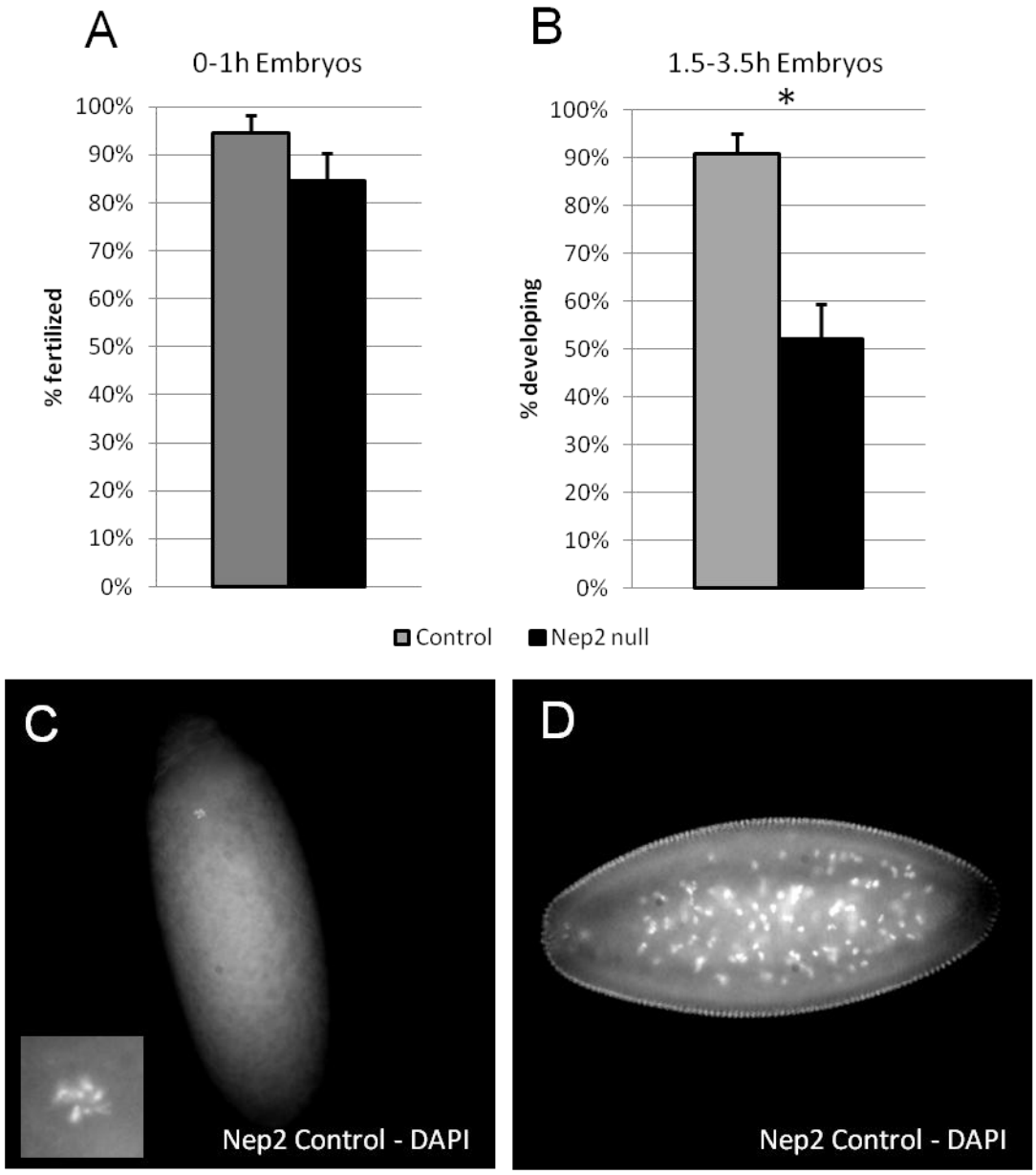


Figure 2.8 Eggs laid by Nep2 null females arrest during early embryogenesis

A) Eggs laid by Nep2 null females are fertilized at the same rate as eggs laid by control females (WRST p= 0.1593, control N = 37, *Nep2* null N= 39) based on sperm tail staining. **B)** DAPI staining of 1.5-3.5h old eggs laid by Nep2 control or *Nep2* null females were sorted into two categories: developing or non-developing. All non-developing embryos contained a polar body rosette (C), whereas developing embryos were all at stage 4+ (D) of development consistent with the time point chosen. Eggs laid by *Nep2* null females are significantly more likely to fall into the non-developing category than eggs laid by control females (WRST p < 0.0001*, Control N= 43, *Nep2* null N=48). Since the fertilization rate between Nep2 null and control females is not different this result suggests that *Nep2* may be critical for early embryogenesis.

activated but unfertilized eggs (Figure 2.8C) whereas the typical developing embryo at this time range was observed to be at stage 4 (Figure 2.8D). Since there is no difference in fertilization rate between eggs laid by *Nep2* null females and controls, these data suggest that some of the eggs laid by *Nep2* null females are able to activate and complete meiosis but fail to develop further. This fraction of arrested eggs observed for *Nep2* mutant females is consistent with the magnitude of the hatching defects seen above (Fig. 2.7) and suggests that *Nep2* plays a role in egg-laying and has a maternal effect on very early embryogenesis.

Tissue specific contribution to egg-laying and hatchability effects for Nep1 and Nep2 females

The egg-laying defects in *Nep1* and *Nep2* RNAi females suggested that one or more of the tissues in which we detected Nep expression must be essential for egg-laying. Both Nep1 and Nep2 are expressed in the CNS and the spermathecae, two tissues that are known to influence egg-laying [25,31]. Nep2 is also expressed in the border cells of the follicular epithelium, which are important for micropyle development and for anterior-posterior polarity in the egg [32]. To test whether these tissues require Nep function for normal egg-laying or hatchability we individually used *nsyb*-GAL4 [33], *Send1*-GAL4 [25], and *slbo*-GAL4 [34] drivers to locally drive RNAi in the CNS, the spermathecae, and the border cells, respectively.

Females knocked-down for either *Nep1* or *Nep2* in the CNS laid fewer eggs than control females (Figure 2.9 & 2.10). Similarly, knockdown of *Nep1* or *Nep2* in the spermathecae also reduced egg-laying (Figure 2.9 & 2.10). However, knockdown in either tissue did not fully recapitulate the egg-laying phenotypes seen in the ubiquitous knockdown of *Nep1* or *Nep2* in females. Contrary to expectations, reduction of *Nep2* expression in the

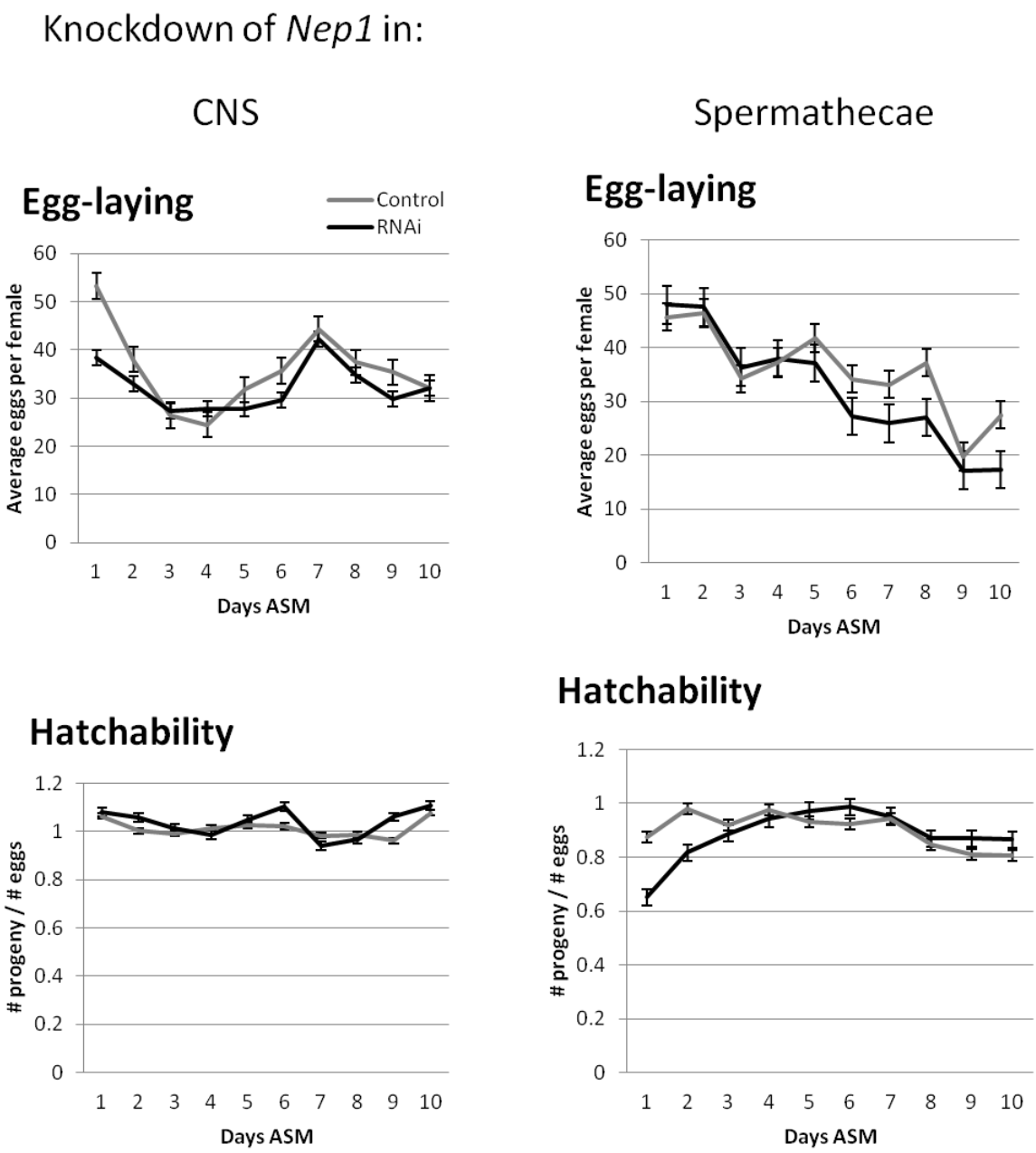


Figure 2.9 Egg-laying in tissues specific knockdowns of *Nep1* in females

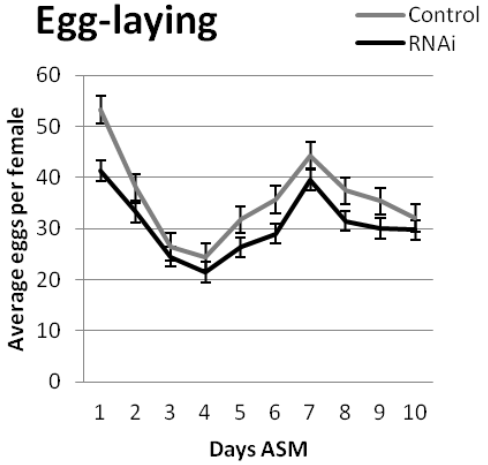
The mean number of eggs laid by Control (grey line) or *Nep1* RNAi females (black line) knocked down in one of two tissues (the CNS or the spermathecae) mated to WT males over a 10 day period. Females that have reduced Nep1 expression in either the CNS (rmANOVA p=0.0042, Control N=18 , RNAi N=18) or spermathecae (rmANOVA p=0.0078, Control N= 17, RNAi N=13) lay fewer eggs than control females. There was no overall effect of reduction in *Nep1* expression on hatchability in the case of the CNS (rmANOVA p=0.0853) or the spermathecae (rmANOVA p=0.4419) although loss of *Nep1* in the spermathecae can reduce initial hatchability on Days 1 and 2 (Day 1 WRST p=0.0161, Day 2 WRST p=0.0005)

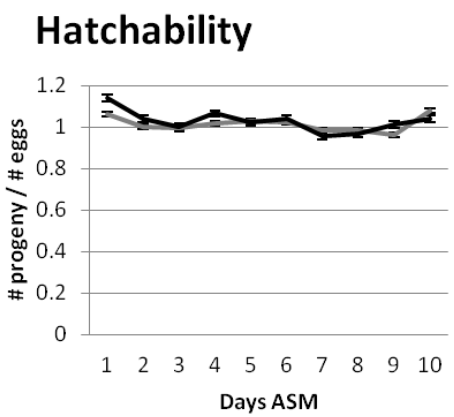
Figure 2.10 Egg-laying in tissues specific knockdowns of *Nep2* in females

The mean number of eggs laid by Control (grey line) or Nep2 RNAi females (black line) knocked down in one of three tissues (the CNS, spermathecae, or border cells) mated to WT males over a 10 day period. Females with reduced Nep2 expression in either the CNS (rmANOVA p=<0.0001*, Control N= 18, RNAi N=16) or spermathecae (rmANOVA p=0.0002, Control N= 19, RNAi N=15) laid fewer eggs than control females, whereas loss of *Nep2* expression in the border cells slightly increased egg-laying (rmANOVA p=0.0177, Control N= 23, RNAi N=17). None of these experiments recapitulate the hatchability defects seen in Nep2 null females (CNS rmANOVA p=0.1221, Border Cells rmANOVA p=0.1170).

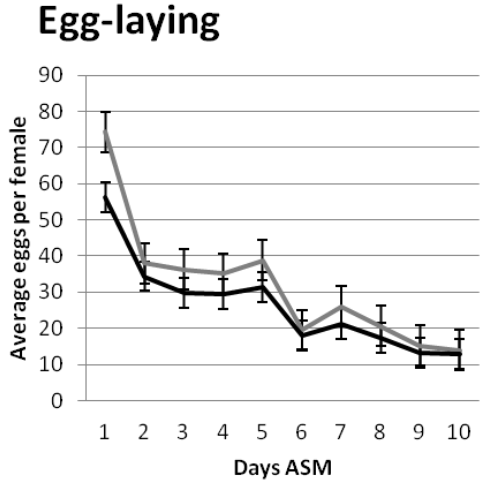
Knockdown of *Nep2* in:

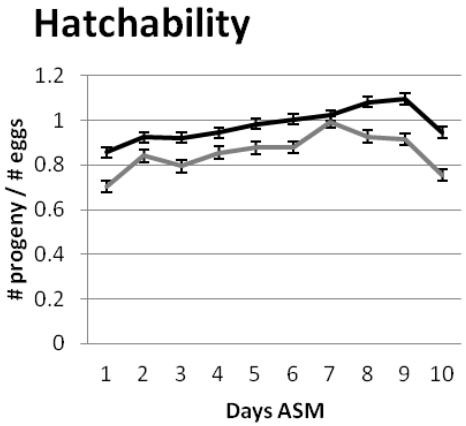
CNS



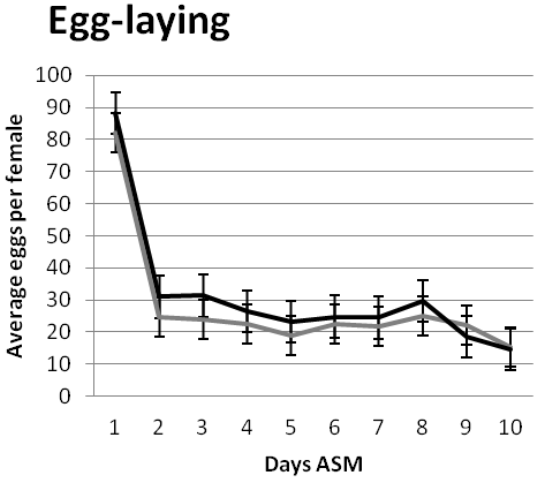


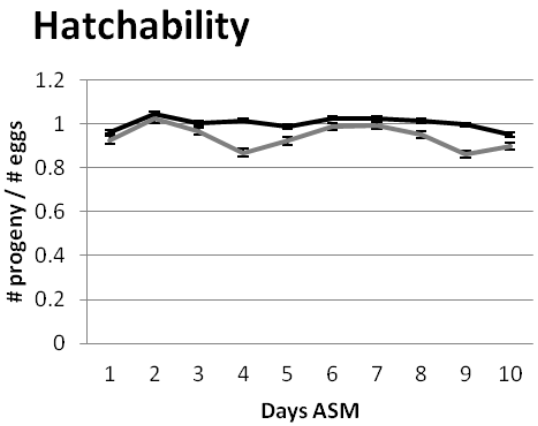
Spermathecae





Border Cells





border cells slightly increased egg-laying. However, none of the targeted tissues for *Nep2* resulted in a decrease in hatchability. These results suggest that Neps are important in both the CNS and the spermathecae for normal egg-laying but not hatchability and that either a combination of both sources or an as yet untested source of *Nep* expression may be responsible for the majority of the reduction in egg-laying observed. One possible source of expression is the seminal receptacle, where *Nep1* and *Nep2* transcripts have both been detected [35]; however, we do not currently have GAL4 drivers that target this organ.

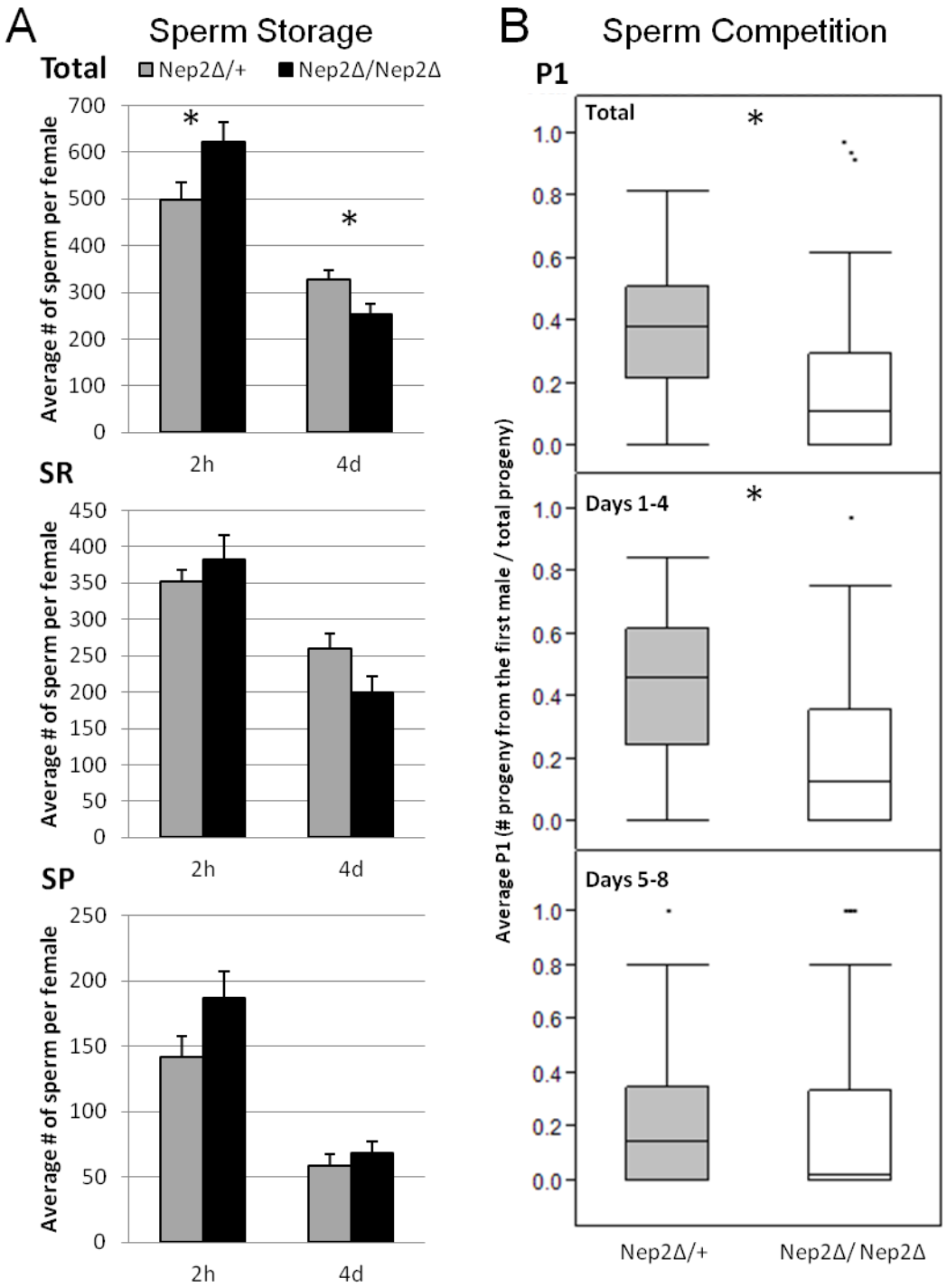
Sperm storage and depletion are abnormal in Nep2 null females

Neprilysins have been implicated in the regulation of muscle contraction in the mammalian uterus [12,13]. Muscle contractions are also important in the *Drosophila* uterus, which goes through conformational changes after mating that facilitate sperm storage [36,37]. The spermathecae and the seminal receptacle, which store these sperm, also experience contractions [38], the importance of which is unknown but may aid in facilitating the release or storage of sperm. To determine whether *Nep2* is essential for sperm storage or release, we counted the number of sperm stored at 2h and 4d ASM in *Nep2* null females compared to controls.

After mating to wild type males, females that were null for *Nep2* stored more sperm overall at 2h post mating (a time when sperm storage has just completed [39]) than control females (Figure 2.11A), and had marginally more sperm in the spermathecae. By 4d ASM, however, *Nep2* null females stored fewer sperm than controls and had marginally fewer sperm stored in their seminal receptacles. This shift from surplus to deficit illustrates that *Nep2* null females are defective in sperm retention. These results suggest that *Nep2* plays a role not only in the initial storage of sperm but also in controlling the release of sperm from the sperm storage organs. Although *Nep2* plays a role in sperm storage and release, the

Figure 2.11 Sperm utilization in *Nep2* null females

A) Counts of sperm stored in both sets of sperm storage organs (Total), the seminal receptacle (SR) and the paired spermatheca (SP), of *Nep2* null (black) versus control females (grey) at 2h and 4d after the start of mating (ASM). Overall Nep2 null females store more sperm at 2h ASM (ANOVA, F= 4.8029. p=0.0398, Control N=13, *Nep2* null N=10) and fewer sperm at 4d ASM (ANOVA, F=6.0175 p=0.0215*, Control N=13, Nep2 null N=14) than control females. Within the SR Nep2 null females store the same number of sperm at 2h ASM (ANOVA, F=0.71 p=0.4061, Control N=17, *Nep2* null N=15) and marginally fewer sperm at 4d ASM (ANOVA, F=3.920 p=0.0580, Control N=14, *Nep2* null N=15) than controls. Within the SP *Nep2* null females store the same number of sperm at both 2h ASM (ANOVA, F=3.1304. p=0.0901, Control N=13, *Nep2* null N=12) and 4d ASM (ANOVA, F=0.5584 p=0.4614, Control N=14, Nep2 null N=15). **B)** For sperm competition assays *Nep2* null or control females were first mated to a Canton-S male and then allowed to mate a second time with a *bwD* male. The proportion of female progeny sired by the first male (Canton-S) referred to as P1(# progeny from first male / total progeny) was significantly reduced in Nep2 null females compared to control females (WRST p=<0.0001* , Control N=76, *Nep2* null N=72). This difference is most apparent in the first 4d ASM (WRST p=<0.0001*) compared to days 5-8 (WRST p=0.1886)


A
Sperm Storage
Total
Nep2Δ/+
Nep2Δ/Nep2Δ
Average # of sperm per female
700
600
500
400
300
200
100
0
*
*
2h
4d
SR
450
400
350
300
250
200
150
100
50
0
2h
4d
SP
250
200
150
100
50
0
2h
4d
B
Sperm Competition
P1
Total
Days 1-4
Days 5-8
Average P1 (# progeny from the first male / total progeny)
1.0
0.8
0.6
0.4
0.2
0.0
*
*
Nep2Δ/+
Nep2Δ/ Nep2Δ

number of sperm stored in the sperm storage organs at the 4d time point is too high to suggest that these sperm storage differences alone underlie the egg-laying defects seen in *Nep2* null females.

To confirm the role of *Nep2* in regulating sperm release or depletion we performed a sperm competition assay in which we mated *Nep2* null and control females to a Canton-S male and then subsequently to a bw^D male. Loss of *Nep2* function dramatically decreases P1 (the proportion of progeny sired by the first male) (Figure 2.11B) suggesting that *Nep2* aids in sperm retention and works to help sperm resist displacement by rival ejaculates. This is consistent with the observation that sperm deplete faster in singly mated Nep2 null females. Together these results indicate a role for Nep2 in female regulated sperm use.

2.3. DISCUSSION

Drosophila neprilysin genes

We investigated a group of genes encoding M13 class proteases in *Drosophila melanogaster* with expression patterns suggesting that they may play roles in reproduction or the CNS. Sequence analysis of protein sequences of Nep1-Nep5 with the sequences of human family members ECE-1, ECE-2 and neprilysin and locust *Lom*ECE and phylogenetic analysis revealed distinct similarities for *Drosophila* neprilysins 1-5. *Drosophila* Nep1 and Nep4 are most closely related to a group of vertebrate neprilysin homologues. Nep3 is most similar to *Lom*ECE and vertebrate ECE and Nep5 clusters in a group with Kell homologs. Nep2 is an invertebrate specific protein. Overall, our analysis indicates that Nep1-5 are evolutionarily closely related yet representative of the functional divergence that seems to have occurred in this gene family.

Implications of Neprilysins 1-5 expression patterns.

The strong conservation of domains important for correct protein folding and activity in Nep1-Nep5 suggests that the functional specificity of the enzymes may at least in part depend on their specific spatiotemporal expression patterns, an aspect that has previously also been observed in *C. elegans [1]*. Analysis of the expression patterns of *Nep1-Nep5* by *in situ* hybridization supports this hypothesis. Although the major role of these enzymes is proposed to be the metabolism of neuropeptides and peptide hormones, the presence and role of (neuro)peptides in the male and female reproductive organs is currently unknown. Overall our expression analysis suggests that these five neprilysins may be involved in a range of developmental and physiological processes that in turn may be mediated by numerous bioactive (neuro)peptides.

Drosophila ***Neprilysins are important for fertility***

Both *Nep1* and *Nep2* are essential for normal female reproductive fitness. *Nep1* is essential for egg-laying and may contribute to egg production. Part of the effects of *Nep1* expression on egg-laying can be traced to its role in the spermathecae and the CNS. *Nep2* is also essential for both the post-mating increase in female egg-laying and the hatchability of laid eggs. The hatching defects seen in *Nep2* RNAi and null females is not due to a failure in fertilization but instead manifests in an early embryonic arrest, suggesting that maternal *Nep2* is essential early in development. Expression of *Nep2* in both the CNS and the spermathecae contributes to the egg-laying defect but not to the hatchability defect. Surprisingly, even though *Nep2* is present in the border cells of the follicular epithelium, the expression of *Nep2* in these cells is not essential for fertility.

In addition to egg production, *Nep2* also influences sperm storage and depletion in females. Loss of *Nep2* in the female also negatively impacts retention of sperm from the first

mating when a second mating occurs. This suggests that *Nep2* may play a role in sperm retention, helping to insulate stored sperm from displacement by rival ejaculates. Whether or not this reduction in the presence of the first male's sperm is detrimental to the female is unclear. Together our data paint a broad role for Neprilysins, and particularly for insect-specific *Neps* like *Nep2*, in regulating female reproductive success. Nep1 is also important in male reproductive fitness. Knockdown of *Nep1* in males decreased egg-laying in their mates. This finding is consistent with experiments in mice, where loss of NL1 in males caused reduced litter sizes [11]. In contrast *Nep2-5* do not appear to have nonredundant, essential roles in male fertility.

The Neps we tested represent only a fraction of the neprilysin-like homologs identified to date in *Drosophila*. The similarity between Nep1 and other vertebrate Neps makes it a good potential model for finding substrates for neprilysins that are conserved throughout female reproduction. Whereas *Nep2* offers insight into the insect specific lineages of neprilysin-like genes. Further research on the substrates of Nep2 may reveal divergent or species-specific mechanisms for neprilysins in reproduction. Substrates of Nep2, or Nep2 itself, could also prove to be useful targets for controlling pests and insect disease vectors by reducing fertility.

Acknowledgements

We would like to acknowledge the support of KULeuven and VIB to P.C. and grants RO1-038921 to M.F.W. and RO1-HD059060 to M.F.W. and Andrew Clark for support of this work. We would also like to thank Daniel Barbash for the use of his Leica CTR5000 microscope, Amber Krauchunas and Norene Buehner for their advice and assistance in embryo staining, Geoff Findlay for comments on this manuscript and his help with MUSCLE,

and the members of the Cornell Superfly Club for helpful suggestions and comments on this manuscript.

2.4. MATERIALS AND METHODS

Sequence comparison and tree building

Protein sequences were downloaded from Flybase [40] and aligned using Muscle [41], and the alignment was checked by eye in MEGA 5.05 [42]. The program ProML, part of the Phylip 3.69 suite, was used to make the tree [43] and it was visualized for publication using FigTree v.1.3.1 [44].

In situ hybridization

Fly culture

yw and Canton-S stocks were maintained on a standard diet (6.4% cornmeal, 5.2% molasses, 1.8% dextrose, 1.2% yeast, 1% propionic acid, 0.75% agar, 0.15% methyl-4-hydroxybenzoate in 1.5% ethanol) at 25°C in plastic vials.

Egg collection

Flies were allowed to lay eggs during 17 hours on apple juice agar plates (3% agar, 5.5% sucrose, 2.5% EtOH, 1.25% glacial acetic acid in apple juice) supplemented with yeast paste in a 25°C incubator. Adult flies were then removed, and the embryos were washed off the original plates with water and transferred to a nylon mesh. To remove remainders of the yeast paste and apple juice agar the embryos were washed with water. Embryos were dechorionated, permeabilized and fixed as described [45].

In situ hybridization: DIG RNA labelling

We used cDNA clones GH03315 (*Nep1-RB*), GH07643 (*Nep2*), RE48040 (*Nep3*), LD25753 (*Nep4*), and AT14086 (*Nep5*), from the *Drosophila* Genomics Research Center (DGRC), for preparation of probes. Overnight restriction digest at 37°C was done with *Not*I and *Bst*BI for antisense and sense *Nep1* probe, *Eco*RI and *Xho*I for *Nep2*, *Not*I and *Asp*718I for *Nep3*, *Eco*RI and *Xho*I for *Nep4*, and *Sal*I and *Nru* I for *Nep5*.

Linearized template DNA was purified using QiaQuick PCR purification kit (Qiagen). RNA labelling was done with the DIG RNA labelling kit (Roche) and 1µg of purified DNA following the manufacturer's protocol. Probes were hydrolyzed to a desired length of 200 bases. The RNA transcripts were analyzed for size by formaldehyde agarose gel electrophoresis and ethidium bromide staining. The labelling efficiency was tested using DIG quantification test strips and control strips (Roche).

Tissue collection, fixation and hybridization

Third instar larval tissue

The posterior end of third instar larvae was removed with forceps and the larvae were inverted to expose the brain, most of the imaginal discs and parts of the gut and fat body.

Adult tissue

Adult abdomens were removed from the thorax and opened on the ventral side from anterior to posterior in order to expose all the tissues to the solutions. The thorax was separated from head and abdomen and the dorsal side of the cuticle was removed. For *in situ* hybridization on adult brains, the proboscis and part of the cuticle and the air sacs were removed from isolated heads.

All dissected tissues were kept in PBT (PBS, 0.1% Tween 20) on ice for maximum one hour before fixation. Fixation was done on a shaker for 60 min at room temperature in 1ml of 4% paraformaldehyde containing 0.1% sodium deoxycholate.

Embryos

Embryos were collected and fixed as described previously. Before starting the proteinase K treatment, embryos were rehydrated in the following conditions for 10 min each: 25% PBT/75% MeOH; 50% PBT/50% MeOH; 75% PBT/25% MeOH and 100% PBT.

All tissues were rinsed in 1ml PBT and washed 5 x 5 min in 1ml PBT after fixation or rehydration. Different tissues were incubated in a volume of 150µl proteinase K mixture: inverted third instar larvae: 15 µg/ml proteinase K for 2 min at 37°C; adult abdomen: 15 µg/ml proteinase K for 3 min at 37°C; adult thorax: 10µg/ml proteinase K for 2 min at 37°C ; adult brain: 10 µg/ml proteinase K for 2 min at 37°C; whole mount embryos: 40 µg/ml proteinase K for 3 min at RT. Prehybridization, hybridization and detection were as described in [46].

Fertility/fecundity assays and sperm competition.

Fly stocks and Media

All flies were raised at room temperature (23±1°C) in glass bottles on standard yeast-glucose media (cite or otherwise distinguish from the above food recipe). Females were aged 3-5 days from eclosion in groups of 5-12 in glass vials with added yeast. Male flies were aged 3-5 days from eclosion in groups of 10-20 in glass vials on standard yeast-glucose media. The RNAi lines used for *Nep1, Nep2, Nep3, Nep4,* and *Nep5* were all obtained from the Vienna Drosophila RNAi Center (VDRC) [47]. Knockdown of transcripts were confirmed by RT-

PCR [48]. A *Nep2* null allele, *Nep2Δ*, was generated by means of a deletion generator compound element as described in [49]. The starting stock was *yw;;P{wHy}*DG19304. Loss of transcript was confirmed by qRT-PCR.

Each RNAi line was crossed to *tubulin*-GAL4/TM3, Sb; the balancer siblings from each cross (UAS-Nep/TM3, Sb) were used as controls to minimize rearing effects. Controls for the other drivers, *n-syb*-GAL4, *slbo*-GAL4, and *Send1*-GAL4 were generated by crossing the VDRC background line w^{1118} to the driver line. In the case of *Nep1*, whose knockdown was lethal with *tubulin*-GAL4, a *hsp70*-GAL4 (HS-GAL4) line was used instead to drive knockdown. *HS*-GAL4;UAS*Nep1*RNAi (or control) males and females were aged for 3 days prior to heat shock as previously described. For heat shock, flies were moved to vials without food that contained a wet piece of Whatman paper, after which they were placed in a water bath at 37°C for one hour. The heat-shocked flies were allowed to recover at room temperature in vials containing fresh food, and were then mated 12 hours later for all assays in which they were used.

Fertility/fecundity assays

In all assays involving male fertility, we used 3-5 day old Canton-S virgin females. Females were placed singly in glass vials with food and allowed access to an RNAi (or mutant) male or control male. Pairs were watched to confirm that mating had occurred. The male was removed upon dismounting. Assays for the effects of the Neps on female fertility were performed the same way using 3-5 day old Canton-S males as mates for either RNAi (or mutant) or control females.

After mating, individual females were housed on yeast glucose media for 24 hours after which each female was transferred to a fresh vial, and the eggs laid in the previous vial were counted as described in [50]. Comparisons of single day and total egg and progeny production between control and experimental females were performed using a Wilcoxon non-parametric test and statistics comparing the overall 10 day trends were performed using a repeated measures ANOVA. All statistical analysis was performed with the Jmp9 software.

Sperm competition

After mating, *Nep2Δ* or control females were individually housed for 3 days on yeast-glucose media after which each female was allowed access to a single bw^D male for 12 hours. After the bw^D male was removed, the females were transferred individually to fresh vials and allowed to lay eggs for 4 days before being transferred to fresh food vials and allowed to lay eggs for an additional 4 days. Because the *Nep2Δ* stock is in a *y w* background and the dominant bw^D eye color phenotype (brown) requires the presence of a *w+* allele to be scored, only female progeny who carried the *w+* allele from the male were scored for the presence of bw^D (provided by the second male) or red eyes (provided by the first male). P1 was calculated as (# progeny sired by the first male) / (total progeny). Comparisons between the P1 of control and experimental females were performed using a one-way ANOVA and by Wilcoxon non-parametric tests.

Embryo collection and staining for development and sperm tails

For assaying the ability for eggs laid by *Nep2Δ* females to develop into embryos, we collected 1.5-3.5 hour old eggs, fixed them using methanol/heptane, and stained with DAPI, as described in [51]. For DAPI staining, fixed embryos were incubated in PBS containing 1 μ /ml DAPI for 5 min and were washed 5 x 15 min in PBST. To assess the presence of sperm

tails in eggs laid by *Nep2Δ* and control females we collected eggs laid in a 1hr window at room temperature and prepared them as previously described except that Rat anti-sperm tail antibody was used at a dilution of 1:800 instead ([30]; T. Karr, ASU, personal communication). Images were collected using a Leica CTR5000 microscope (DAPI) (courtesy of Dan Barbash) or a Leica TCS SP2 confocal microscope (Sperm tail).

Sperm counts

Nep2Δ or control mated females were frozen in liquid nitrogen at 2h ASM or kept in glass vials on yeast-glucose media for 4 days and then frozen. Frozen females were stored at -80^{o}C for less than 2 weeks before counting. Reproductive tracts were dissected and then stained with orcein [52,53,54] A transillumination microscope was used at 1000x magnification to visualize sperm. Comparisons between the number of sperm present in control and experimental females were performed using Wilcoxon non-parametric tests.

4.5 REFERENCES

1. Turner AJ, Isaac RE, Coates D (2001) The neprilysin (NEP) family of zinc metalloendopeptidases: genomics and function. Bioessays 23: 261-269.

2. Turner AJ, Brown CD, Carson JA, Barnes K (2000) The neprilysin family in health and disease. Adv Exp Med Biol 477: 229-240.

3. Bland ND, Pinney JW, Thomas JE, Turner AJ, Isaac RE (2008) Bioinformatic analysis of the neprilysin (M13) family of peptidases reveals complex evolutionary and functional relationships. BMC Evol Biol 8: 16.

4. Segura J, Ruilope LM (2011) Dual-acting angiotensin receptor-neprilysin inhibition. Curr Hypertens Rep 13: 74-78.

5. Wick MJ, Buesing EJ, Wehling CA, Loomis ZL, Cool CD, et al. (2011) Decreased neprilysin and pulmonary vascular remodeling in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 183: 330-340.

6. Klein C, Patte-Mensah C, Taleb O, Bourguignon JJ, Schmitt M, et al. (2013) The neuroprotector kynurenic acid increases neuronal cell survival through neprilysin induction. Neuropharmacology.

7. Mulder SD, Veerhuis R, Blankenstein MA, Nielsen HM (2012) The effect of amyloid associated proteins on the expression of genes involved in amyloid-beta clearance by adult human astrocytes. Exp Neurol 233: 373-379.

8. Wong SS, Sun NN, Fastje CD, Witten ML, Lantz RC, et al. (2011) Role of neprilysin in airway inflammation induced by diesel exhaust emissions. Res Rep Health Eff Inst: 3-40.

9. Maguer-Satta V, Besancon R, Bachelard-Cascales E (2011) Concise review: neutral endopeptidase (CD10): a multifaceted environment actor in stem cells, physiological mechanisms, and cancer. Stem Cells 29: 389-396.

10. Smollich M, Gotte M, Yip GW, Yong ES, Kersting C, et al. (2007) On the role of endothelin-converting enzyme-1 (ECE-1) and neprilysin in human breast cancer. Breast Cancer Res Treat 106: 361-369.

11. Carpentier M, Guillemette C, Bailey JL, Boileau G, Jeannotte L, et al. (2004) Reduced fertility in male mice deficient in the zinc metallopeptidase NL1. Mol Cell Biol 24: 4428-4437.

12. Pintado CO, Pinto FM, Pennefather JN, Hidalgo A, Baamonde A, et al. (2003) A role for tachykinins in female mouse and rat reproductive function. Biol Reprod 69: 940-946.

13. Pinto FM, Armesto CP, Magraner J, Trujillo M, Martin JD, et al. (1999) Tachykinin receptor and neutral endopeptidase gene expression in the rat uterus: characterization

and regulation in response to ovarian steroid treatment. Endocrinology 140: 2526-2532.

14. Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39: 715-720.

15. Coates D, Siviter R, Isaac RE (2000) Exploring the Caenorhabditis elegans and Drosophila melanogaster genomes to understand neuropeptide and peptidase function. Biochem Soc Trans 28: 464-469.

16. Isaac RE, Parkin ET, Keen JN, Nassel DR, Siviter RJ, et al. (2002) Inactivation of a tachykinin-related peptide: identification of four neuropeptide-degrading enzymes in neuronal membranes of insects from four different orders. Peptides 23: 725-733.

17. Wilson CL, Shirras AD, Isaac RE (2002) Extracellular peptidases of imaginal discs of Drosophila melanogaster. Peptides 23: 2007-2014.

18. Bland ND, Thomas JE, Audsley N, Shirras AD, Turner AJ, et al. (2007) Expression of NEP2, a soluble neprilysin-like endopeptidase, during embryogenesis in Drosophila melanogaster. Peptides 28: 127-135.

19. Meyer H, Panz M, Zmojdzian M, Jagla K, Paululat A (2009) Neprilysin 4, a novel endopeptidase from Drosophila melanogaster, displays distinct substrate specificities and exceptional solubility states. J Exp Biol 212: 3673-3683.

20. Thomas JE, Rylett CM, Carhan A, Bland ND, Bingham RJ, et al. (2005) Drosophila melanogaster NEP2 is a new soluble member of the neprilysin family of endopeptidases with implications for reproduction and renal function. Biochem J 386: 357-366.

21. Rawlings ND, Barrett AJ, Bateman A (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 40: D343-350.

22. Li C, Booze RM, Hersh LB (1995) Tissue-specific expression of rat neutral endopeptidase (neprilysin) mRNAs. J Biol Chem 270: 5723-5728.

23. Ouimet T, Facchinetti P, Rose C, Bonhomme MC, Gros C, et al. (2000) Neprilysin II: A putative novel metalloprotease and its isoforms in CNS and testis. Biochem Biophys Res Commun 271: 565-570.

24. Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, et al. (2009) Unlocking the secrets of the genome. Nature 459: 927-930.

25. Schnakenberg SL, Matias WR, Siegal ML (2011) Sperm-storage defects and live birth in Drosophila females lacking spermathecal secretory cells. PLoS Biol 9: e1001192.

26. Dow JA, Romero MF (2010) Drosophila provides rapid modeling of renal development, function, and disease. Am J Physiol Renal Physiol 299: F1237-1244.

27. Brand AH, Perrimon N (1993) Targeted Gene-Expression as a Means of Altering Cell Fates and Generating Dominant Phenotypes. Development 118: 401-415.

28. Halfon MS, Kose H, Chiba A, Keshishian H (1997) Targeted gene expression without a tissue-specific promoter: creating mosaic embryos using laser-induced single-cell heat shock. Proc Natl Acad Sci U S A 94: 6255-6260.

29. Silbermann R, Tatar M (2000) Reproductive costs of heat shock protein in transgenic Drosophila melanogaster. Evolution 54: 2038-2045.

30. Karr TL (1991) Intracellular sperm/egg interactions in Drosophila: a three-dimensional structural analysis of a paternal product in the developing egg. Mech Dev 34: 101-111.

31. Yang CH, Rumpf S, Xiang Y, Gordon MD, Song W, et al. (2009) Control of the postmating behavioral switch in Drosophila females by internal sensory neurons. Neuron 61: 519-526.

32. Furriols M, Ventura G, Casanova J (2007) Two distinct but convergent groups of cells trigger Torso receptor tyrosine kinase activation by independently expressing torso-like. Proceedings of the National Academy of Sciences of the United States of America 104: 11660-11665.

33. Pauli A, Althoff F, Oliveira RA, Heidmann S, Schuldiner O, et al. (2008) Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. Dev Cell 14: 239-251.

34. Rorth P, Szabo K, Bailey A, Laverty T, Rehm J, et al. (1998) Systematic gain-of-function genetics in Drosophila. Development 125: 1049-1057.

35. Prokupek AM, Eyun SI, Ko L, Moriyama EN, Harshman LG (2010) Molecular evolutionary analysis of seminal receptacle sperm storage organ genes of Drosophila melanogaster. J Evol Biol 23: 1386-1398.

36. Adams EM, Wolfner MF (2007) Seminal proteins but not sperm induce morphological changes in the Drosophila melanogaster female reproductive tract during sperm storage. J Insect Physiol 53: 319-331.

37. Avila FW, Wolfner MF (2009) Acp36DE is required for uterine conformational changes in mated Drosophila females. Proc Natl Acad Sci U S A 106: 15796-15800.

38. Middleton CA, Nongthomba U, Parry K, Sweeney ST, Sparrow JC, et al. (2006) Neuromuscular organization and aminergic modulation of contractions in the Drosophila ovary. BMC Biol 4: 17.

39. Bloch Qazi MC, Heifetz Y, Wolfner MF (2003) The developments between gametogenesis and fertilization: ovulation and female sperm storage in Drosophila melanogaster. Dev Biol 256: 195-211.

40. Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J, et al. (2013) FlyBase: improvements to the bibliography. Nucleic Acids Res 41: D751-757.

41. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

42. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.

43. Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.: Distributed by the author.

44. Rambaut A (2010) FigTree 1.3.1

45. Sullivan W, Ashburner, M., Hawley, R., editor (2000) Drosophila Protocols. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

46. Clements J, Hens K, Francis C, Schellens A, Callaerts P (2008) Conserved role for the Drosophila Pax6 homolog Eyeless in differentiation and function of insulin-producing neurons. Proc Natl Acad Sci U S A 105: 16183-16188.

47. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, et al. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448: 151-156.

48. Ram KR, Wolfner MF (2007) Sustained post-mating response in Drosophila melanogaster requires multiple seminal fluid proteins. PLoS Genet 3: e238.

49. Huet F, Lu JT, Myrick KV, Baugh LR, Crosby MA, et al. (2002) A deletion-generator compound element allows deletion saturation analysis for genomewide phenotypic annotation. Proc Natl Acad Sci U S A 99: 9948-9953.

50. Gligorov D, Sitnik JL, Maeda RK, Wolfner MF, Karch F (2013) A novel function for the hox gene abd-B in the male accessory gland regulates the long-term female post-mating response in Drosophila. PLoS Genet 9: e1003395.

51. Krauchunas AR, Horner VL, Wolfner MF (2012) Protein phosphorylation changes reveal new candidates in the regulation of egg activation and early embryogenesis in D. melanogaster. Dev Biol 370: 125-134.

52. Avila FW, Ravi Ram K, Bloch Qazi MC, Wolfner MF (2010) Sex peptide is required for the efficient release of stored sperm in mated Drosophila females. Genetics 186: 595-600.

53. Neubaum DM, Wolfner MF (1999) Mated Drosophila melanogaster females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153: 845-857.

54. Mueller JL, Linklater JR, Ravi Ram K, Chapman T, Wolfner MF (2008) Targeted gene deletion and phenotypic analysis of the Drosophila melanogaster seminal fluid protease inhibitor Acp62F. Genetics 178: 1605-1614.

CHAPTER 3

MOLECULAR CHARACTERIZATION AND EVOLUTION OF A GENE FAMILY ENCODING BOTH FEMALE- AND MALE-SPECIFIC REPRODUCTIVE PROTEINS IN *DROSOPHILA*[3]

3.1 INTRODUCTION

Gene duplication provides the opportunity for the evolution of novel protein functions through modifications to the original gene or its new paralog. These new functions could result from changes to protein sequences resulting from mutations in the protein-coding region of one of the paralogs or modifications to transcript splicing patterns [1,2,3,4,5]. New functions also could result from mutations in regulatory regions that cause changes in the tissue or timing of gene expression, or of the timing or level of the protein's synthesis [2,6,7,8], or from differential susceptibility to epigenetic modifications [9]. Such changes in the expression of paralogs can have profound impacts on developmental patterns (e.g., Hox genes; [10]). They have also been proposed to provide a resolution to intralocus sexual conflict through the evolution of differential expression patterns in the two sexes [11,12,13,14,15].

Across a range of taxa and timescales, gene duplication events have played major roles in shaping each species' suite of reproductive proteins. In abalone (genus *Haliotis*), for example, extant species have two major acrosomal proteins with distinct functions: lysin, which dissolves a hole in the vitelline envelope surrounding the egg, and Sp18, which mediates sperm-egg fusion

[3]A version of this chapter will be submitted for publication as " Sirot, L.K.*, Findlay, G.F.*, Sitnik, J.L.*, Frasheri, D., Avila, F.W., Wong, A., and Wolfner, M.F., "Molecular Characterization and Evolution of a Gene Family Encoding Both Female- and Male-Specific Reproductive Proteins in *Drosophila"* (* contributed equally). Figures and data contributed by other authors are denoted in the figure legends. My main contribution to this project was all of the phenotyping data. Supplemental material is not included unless produced by J. Sitnik or essential to the paper.

[16,17]. Both proteins have evolved rapidly, to the point where their primary amino acid sequences cannot be aligned, but their similar sizes, molecular weights, three-dimensional structures and patterns of exons and introns suggest an ancient duplication event of a single protein, presumably followed by sub-functionalization [18,19]. In insects, an ancient duplication pre-dating the divergence of many orders gave rise to several isoforms of tubulin, including the testis-specific isoform, β2 [20,21]. This isoform is essential for sperm axoneme function and is thus highly conserved among those species that share it [22,23]. Subsequent to the duplication that gave rise to the original β2, specific insect lineages (including wasps, bees, pea aphids and stalk-eyed flies) experienced duplications of β2, and the new paralogs have evolved more rapidly than the conserved β2 [21,24]. Additional examples of reproductive gene duplication, which are often followed by divergence driven by positive selection, are seen in mammalian Nod-like receptors expressed in the ovary [25], a set of tandemly duplicated serine proteases expressed in female reproductive tissues in *Anopheles gambiae* [26], and a variety of testis-expressed *Drosophila* genes [27,28,29].

Functional consequences of reproductive protein duplication have been studied in several cases involving *Drosophila*. In one example, a male-expressed gene, *Gld2*, arose from the duplication of the ancestral *wispy* (*wisp*) gene that is found in *Drosophila*, *Caenorhabditis* and *Xenopus* [30]. The WISP and GLD2 proteins are each cytoplasmic regulators of messenger RNA poly(A) tail stability, but whereas WISP (and its worm and frog orthologs) act in the female germline [30], GLD2 is instead expressed specifically in the male germline and regulates mRNAs in spermatogenesis. Each protein plays an essential role in reproduction: WISP is required for female fertility (playing essential roles in oogenesis and egg activation; [31,32]), and GLD2 is necessary for male fertility (playing essential roles in the production of mature sperm;

[30]). Duplication events may also provide evidence for co-evolution of reproductive proteins between the sexes. For example, in a desert *Drosophila* species, *D. arizonae*, female reproductive tracts express numerous, functional paralogs of several proteases [33], while males produce seminal fluid that contains several duplicated protease inhibitors [34]. Proteolytic enzymatic activity in female reproductive tracts decreases upon mating, suggesting interactions between proteases in the female and inhibitors from the male [35].

Gene duplication has been an important force for generating diversity among seminal fluid proteins (Sfps), a class of reproductive proteins that exerts dramatic effects on female physiology and behavior [36,37]. In rodents, several duplication events gave rise to the family of seminal vesicle secretion (Svs) genes [38,39]; together, their proteins comprise a substantial fraction of the copulatory plug [40]. Interestingly, primate homologs of these genes have recently undergone further duplication and diversification [41]. In several species of *Anopheles* mosquitoes, duplications have generated three paralogs (termed *AgAcp334A1-3*) of an Sfp hypothesized to regulate sperm motility, and at least one of these proteins is found in the copulatory plug in mated *A. gambiae* females [42,43]. *Drosophila* seminal fluid is replete with proteins encoded by gene duplicates. Proteomic analysis in *D. melanogaster* found that more than 30 out of 133 transferred Sfps were encoded by a gene with an apparent tandem paralog that also encoded a transferred Sfp [44]. Pervasive Sfp duplication is also seen in the *repleta* group of *Drosophila*, but with an intriguing difference: while *D. melanogaster* group duplicates tend to be highly diverged and more ancient in origin (i.e., duplication likely pre-dates the divergence of *D. melanogaster* from its closest sibling species), there are several cases of recently duplicated, adaptively evolving seminal protein pairs between the closely related species, *D. arizonae* and *D. mojavensis* [45]. Furthermore, the gene that encodes the Sfp known as sex peptide (SP) has been

tandemly duplicated in *D. subobscura* [46], though the functional impact of this duplication remains unknown. Thus, duplication of existing Sfp-encoding genes is an important source of Sfp diversity across taxa.

Gene duplication also could act to incorporate new types of proteins into seminal fluid through the duplication and subsequent change in expression pattern of non-Sfp encoding genes (similar to the case of β2 tubulin described above). Evidence for such reproductive "co-option" of proteins comes from observations that specific members of large gene families, most members of which function outside of reproduction, are sometimes found in the seminal fluid. Examples in flies come from the detection in seminal fluid [44] of members of families of serine proteases and protease homologs [47], odorant binding proteins [48,49], and acid lipases [50]. In these cases, the members of each family that are specific to reproduction are found in specific clades or single lineages on the protein phylogenetic tree of the family, which are surrounded by non-reproductive family members. This observation suggests that the reproductive family members gained that pattern of expression relatively recently. Specific members of these classes of proteins have also been found in the seminal fluid of other insects [42,51,52] and of mammals (e.g., [25,38]).

A special case of an Sfp arising from the duplication of a non-Sfp could occur through the duplication and subsequent change in expression pattern of a gene encoding a secreted female reproductive protein such that it is now expressed in the male reproductive tract (analogous to the case of WISP/GLD2 in *Drosophila* gonads, described above). Such a mechanism, while not previously documented to our knowledge, could potentially be an evolutionarily rapid means of creating an effective Sfp: the newly-derived Sfp would already have functionality in the female and, because of its secretion signal sequence, would already be a

prime candidate for transfer from male to female at mating. To search for such a case, we screened 20 secreted, female-specific reproductive proteins in *D. melanogaster* for paralogs that are known Sfps. We report the first case of the evolution of a novel seminal fluid protein gene through the duplication of a gene that has highly biased expression in the female tract. We examine the evolution and expression patterns of this gene family across *Drosophila* species and find evidence for at least two rounds of gene duplication, followed, in one case, by a switch in expression from the female to the male reproductive tract. Each of the five species in the *melanogaster* subgroup has three paralogs, which have maintained consistent patterns of expression. RNA interference testing of the genes singly and in combination in *D. melanogaster* suggests that members of this gene family affect both egg-laying rate and probability of remating.

3.2 RESULTS

A Targeted Search for Female Reproductive Proteins with Sex-switched Duplicates

We found two sets of genes that are highly expressed in the female sperm storage organs and that had an apparent paralog in male seminal fluid. The first set includes the lipase family members, *Yp1*, *Yp2*, and *Yp3*, which are reported to be expressed in the fat body and spermathecae [53] and share highest sequence similarity with each other. These proteins are next most closely related (as measured by sequence similarity) to an Sfp, CG5162. CG5162 shows greatest identity with two other proteins, CG5665 (expressed in embryos; [54]) and CG18258 (expressed in male accessory glands and female spermathecae; [53]), but the Yp proteins are its next most closely related lipase family members [50]. The second set of proteins that fit our criteria includes CG9897 and CG32834, which are reciprocal best BLAST hits and both highly expressed in the

spermathecae. Together, they share highest sequence similarity with CG32833, an Sfp. The three genes encoding these proteins are clustered in a 4-kb region of chromosome 2R. The next-closest gene to this cluster (*CG9896*) is over 7 kb away from the 5' end of *CG9897* and shows no protein sequence similarity to CG9897, CG32834 and CG32833. Because all of the genes in this cluster are located together in the same genomic region, and because the group had no confounding, non-reproductive member (such as *CG5665* in the first case above), we focused on this group of genes (*CG9897*, *CG32833* and *CG32834*) for subsequent evolutionary, expression and functional analysis. Additionally, we found three genes (in bold in Table 1) that are highly expressed in the female sperm storage organs [53] and also encode Sfps [44].

Identification of Family Members in non-melanogaster Species

Putative orthologs were identified based on reciprocal best BLASTP results. We identified three members of this gene family in each of five *Drosophila* species (*simulans*, *sechellia*, *yakuba*, *erecta* and *ananassae*), and one copy in *D. pseudoobscura*. We noticed a fourth paralog, GF11311, annotated in *D. ananassae*, suggesting an extra round of gene duplication in this lineage. However, RT-PCR and sequencing showed that this gene is mis-annotated; as expressed, the gene contains a premature stop codon and thus is likely to be non-functional. The putative orthologs of each *melanogaster* gene are listed in Figure 3.4; as shown below, a combination of gene expression analysis, phylogenetic clustering and conservation of gene order within each syntenic region strongly suggests that these are true orthologs, so we refer to them as such below.

Table 3.1: Genes highly-expressed in *Drosophila melanogaster* female sperm storage organs with information on the presence and identity of seminal fluid protein paralogs.

Family	Gene	Class[a]	SFP paralog	References[b]
Spermathecal endopeptidases (SEND)	CG17012 (SEND1) CG17234 CG17239 CG17240 (ser12) CG18125 (SEND2) CG31861	Serine protease	None	1,2,3,4
Inactive spermathecal endopeptidases (ISEND)	CG9897 CG32834	Inactive serine protease	CG32833	1,3,4
Yolk Protein	CG2985 (YP1) CG2979 (YP2) CG11129 (YP3)	Lipase	CG5162	1,4
Other	**CG6426** [c]	Destabilase	None	1,4
	CG13318	Serine protease	None	3
	CG18067 [c]	3',5'-cyclic-nucleotide phosphodiesterase activity	None	1,4
	CG18525	Serine protease inhibitor	None	1,4
	CG18628 [c]	No conserved domains	None	4
	CG30371	Serine protease	None	4
	CG31686	No conserved domains	None	3
	CG32277	Serine protease	None	1,3,4
	CG32751	Hydrolase	None	3

[a] Based on Flybase.org (Marygold *et al.* 2013), SMART (Letunic et al. 2012), and Pfam classifications (Punta *et al.* 2012).
[b] 1: Allen & Spradling 2008; 2: Arbeitman *et al.*2004; 3: Chintapalli *et al*. 2007; 4: Prokupek *et al*. 2009
[c] Seminal fluid protein-encoding genes (Swanson et al. 2001, Findlay *et al.* 2008)
Table from G. Findlay and L. Sirot.

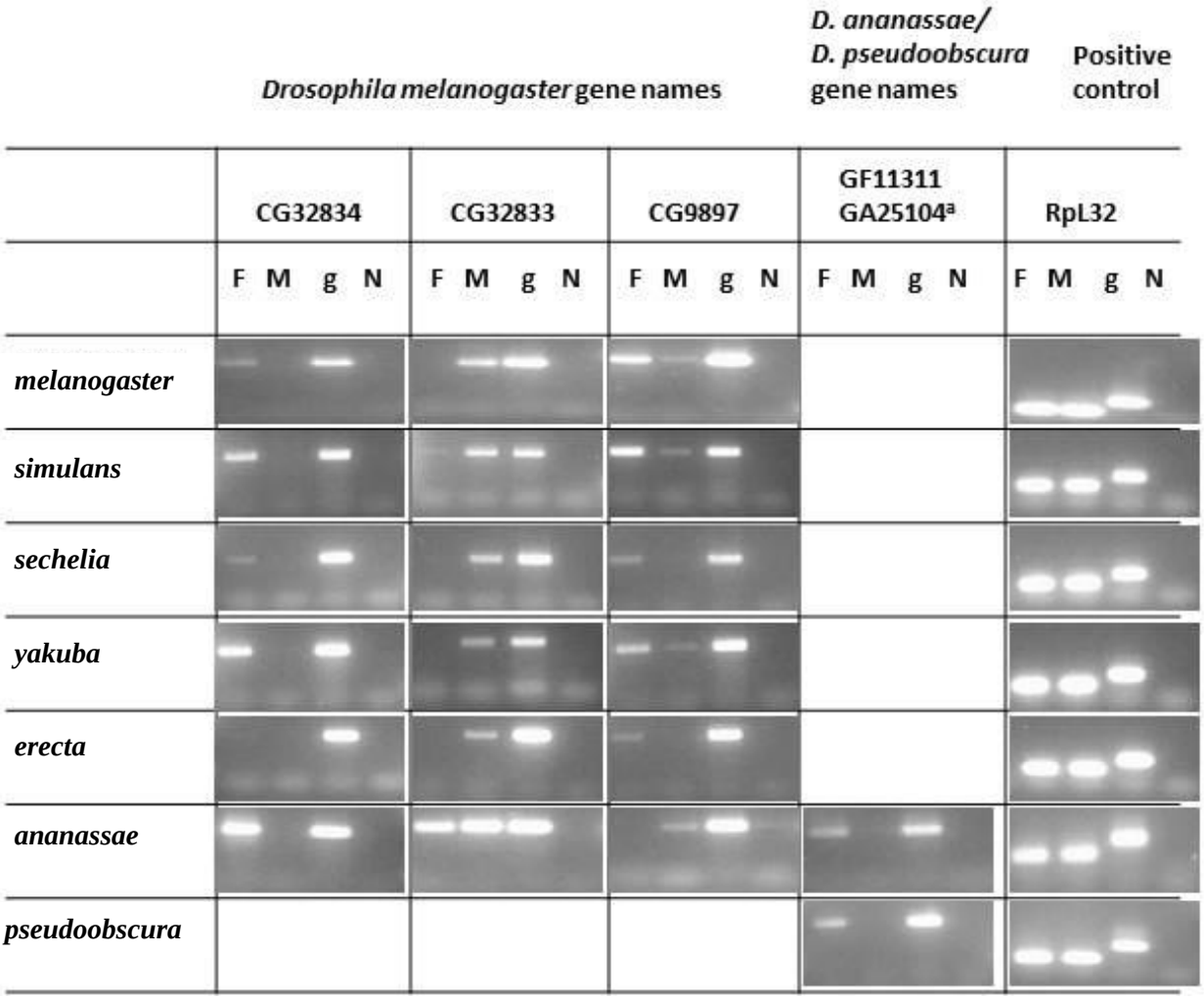
Gene Expression

The gene expression pattern for each of the three *D. melanogaster* genes and almost all of their orthologs is either sex-specific or strongly sex-biased. Most orthologs of *CG32834* and *CG9897* are expressed only in females (Figure 3.1). The exceptions (*CG9897* and its orthologs in *D. yakuba* and *D. erecta*) all have strongly female-biased expression. In contrast, *CG32833* and all of its orthologs in the *melanogaster* subgroup species are expressed only in males, while the *D. ananassae* ortholog, *GF11312*, is expressed in both sexes at apparently equivalent levels. In *D. pseudoobscura*, the single copy of this gene family (*GA25104*) is expressed only in females.

Based on data from microarrays [53], the three *D. melanogaster* genes were thought to be expressed primarily or exclusively in the reproductive accessory glands of males (*CG32833*) or the spermathecae of females (*CG32834* and *CG9897*). We confirmed this gene expression pattern using RT-PCR and further tested for tissue-specific expression patterns of the orthologs found outside of the melanogaster subgroup, in *D. ananassae* and *D. pseudoobscura*. In all cases, expression was either limited to or strongly-biased in the somatic reproductive tissue (reproductive tract without gonads), as compared to the gonads and the carcass (Figure 3.2). Interestingly, while the *D. ananassae* gene *GF11312* shows no sex bias in expression, it is expressed solely in the somatic reproductive tissues of each sex (Figure 3.2).

Phylogenetic Analysis

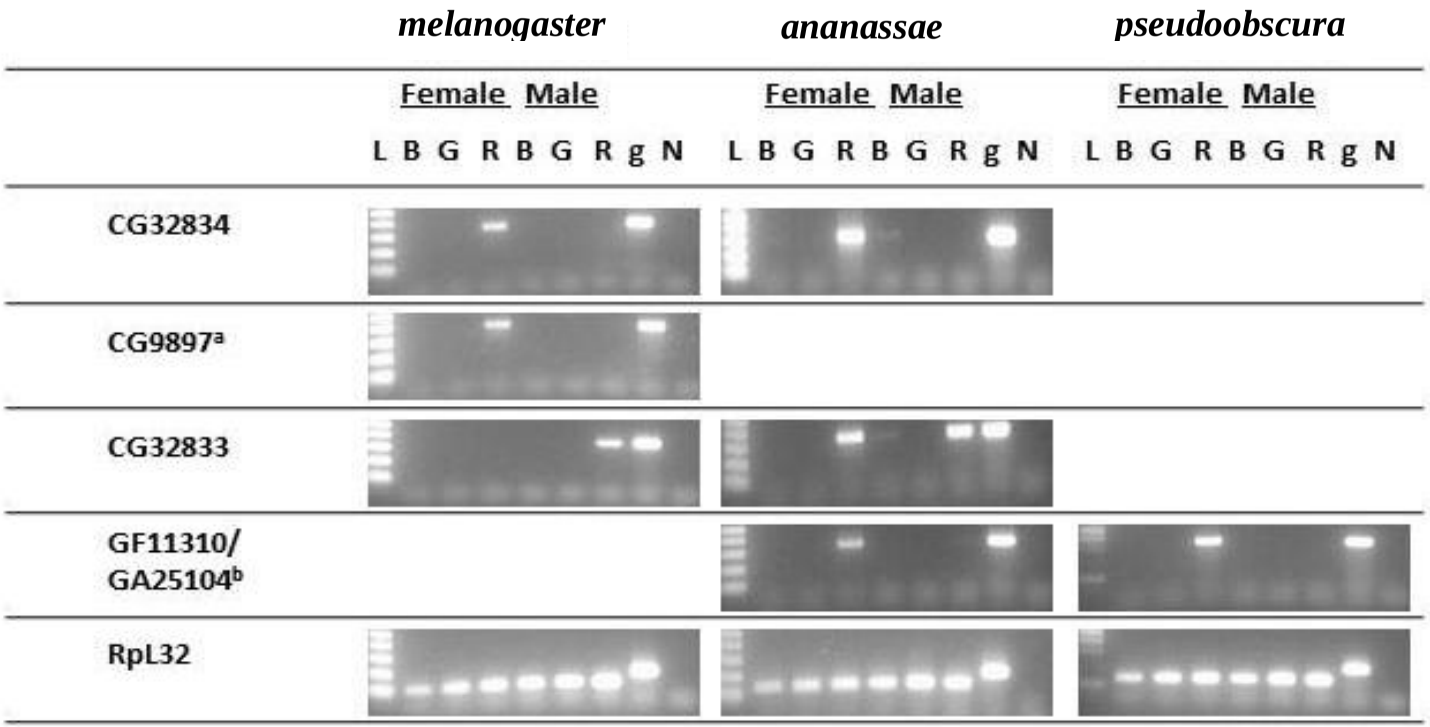
We constructed a protein sequence tree illustrating the degree of amino acid sequence similarity between the core protease domain (corresponding to residues 34-267 of CG32833 of each protein in this family across the melanogaster group of species (*D. melanogaster-D. ananassae*),



[a] GF11310 is a Dana homolog to the gene family but is not homologous with any one gene. GA25104 is a Dpse homolog to the gene family but is not homologous with any one gene.

Figure 3.1: Whole animal expression patterns of *Drosophila melanogaster* CG32834, CG32833, and CG9897 and their orthologs in other congeners

F: female; M: male; G: genomic; N: negative control (water used as template). RpL32 is a ribosomal protein-encoding gene used as a control; primers for this gene were designed to span an intron to check for contamination of cDNA with genomic DNA. Gene names of the orthologs from conspecific species are given in Figure 3.4. Data from D. Frasheri and G. Findlay.



[a]At 35 cycles, a light band appears in the male gonads andreproductive tracts without gonads of Dmel for CG9897
[b]GF11310 is a Dana homolog to the gene family but is not homologous with any one gene. GA25104 is a Dpse homolog to the gene family but is not homologous with any one gene.

Figure 3.2: Tissue specific expression patterns of *Drosophila melanogaster* CG32834, CG32833, and CG9897 and their orthologs in other *D. ananassae*.

Expression patterns are show for GA25104, the single *D. pseudoobscura* ortholog to the gene family. FB: Female body without reproductive tract; FG: Female gonads; FR: Female reproductive tract without gonads; MB: Male body without reproductive tract; MG: Male gonads; MR: Male reproductive tract without gonads; G: Genomic DNA; N: Negative control (water used as template). *RpL32* expression was used as a control. Data from D. Frasheri and G. Findlay

using the single *D. pseudoobscura* copy as the outgroup (Figure 3.3). Phylogenetic clustering was consistent with both the chromosomal order of the genes found in each species and the patterns of expression, supporting our above determinations of orthology. For example, CG32834 and its orthologs are all found at the downstream end of the gene cluster in the genomes of their respective species (Figure 3.4), and all show female-specific/biased expression. The consensus tree formed from 100 bootstrap replicates generally supported the observed tree topology. However, bootstrap support for the most ancestral nodes was low, making it difficult to infer the order of duplication events that gave rise to the extant gene families. We also observed lower bootstrap support for more recent nodes showing the divergence of *D. erecta* and *D. yakuba*, but this result is commonly observed for genes from this pair of species [55]. The more important point is that the tree shows distinct, well-supported clades for each group of orthologs, with the only major ambiguity relating to the assignment of *D. ananassae* GF11310. GF11310 does not cluster with any of the three groups of orthologs, but its female-specific expression pattern and its position at the upstream end of the cluster (Figure 3.4) suggest that it is orthologous to the CG9897 group of genes. The lack of clustering on the tree could be explained by the high level of divergence between GF11310 and other members of the gene family and/or the poor resolution of ancestral nodes.

Phenotypes in Drosophila melanogaster

We examined the phenotypic effects of the three *D. melanogaster* genes using RNAi-mediated knockdown. Specifically, we tested for the effect of knockdown on three post-mating phenotypic responses in females known to be influenced by Sfps [36]: probability of remating, the number of eggs laid, and the number of adult progeny produced.

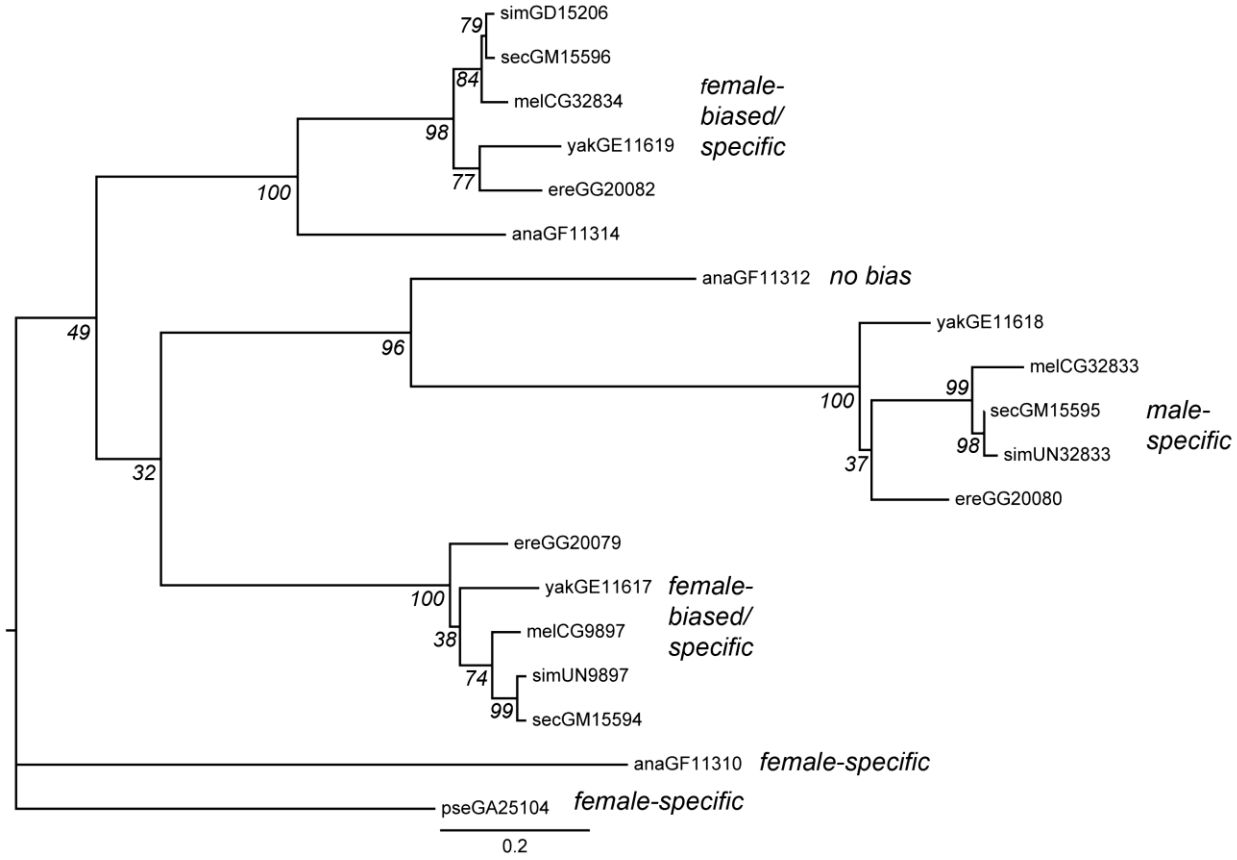


Figure 3.3: Maximum likelihood phylogeny of protein sequences for each member of the gene family.

Bootstrap support based on 100 replicates is shown in italics at each node. Tip labels indicate protein names; the first three letters indicate the *Drosophila* species (mel: *melanogaster*; sim: *simulans*; sec: *sechellia*; yak: *yakuba*; ere: *erecta*; ana: *ananassae*; pse: *pseudoobscura*), and the following characters indicate the FlyBase gene name. "UN" in the gene name indicates a previously unannotated copy of the gene in *D. simulans*. Scale bar indicates the number of substitutions per site. Calls of orthology are consistent with phylogenetic clustering and gene order (see Fig. 4): the six genes shown at the top of the figure (GD15206-GF11314) are one set of orthologs, GF11312-GG20080 are another set, and GG20079-GF11310 are the third set. The tree is rooted on the single *D. pseudoobscura* copy of this gene family, GA25104. Expression patterns from Fig. 1 are indicated in italicized text. Figure from G. Findlay and D. Frasheri.

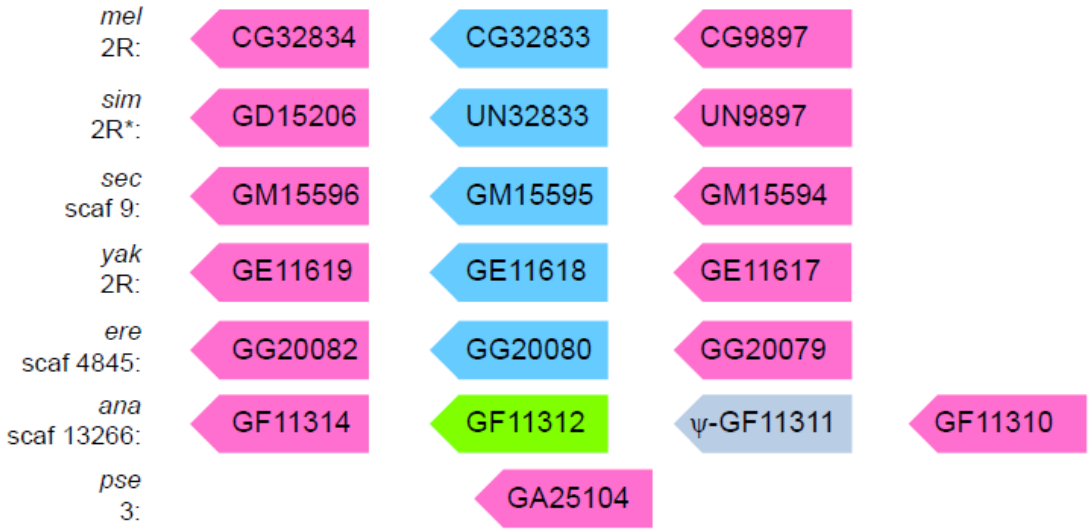


Figure 3.4: Chromosomal locations and gene order of CG9897, CG32833, and CG32834 and their orthlogs in *Drosophila* species.

In *D. simulans*, GD15206 is found in an unassembled part of chromosome (chr) 2R (indicated by the asterisk), while UN32833 and UN9897 represent unannotated copies whose sequences we determined by sequencing or BLAST. The Dsim\UN32833 sequence is only partially determined (the 142 codons at the start of the coding sequence). Color indicates gene expression pattern: pink is female-specific or female-biased; bright blue is male-specific; light blue is expressed only at a low level in males; green is expressed approximately equally in both sexes. Gene order and conserved expression patterns were consistent with calls of orthology and with phylogenetic clustering (Fig. 3). Figure from G. Findlay.

Probability of Remating

We evaluated the effects of knocking down each gene in the cluster on female remating receptivity; results are summarized in Table 3.2 and Figure 3.5 (all χ^2 and *P*-values result from logit loglinear tests). In the first set of experiments (Set 1; Figure 3.5A), we tested for remating after an initial mating between control females and females knocked-down for *CG9897* and *CG32834* individually and in combination mated to either control males or males knocked-down for *CG32833*. Knockdown of individual members of this gene family tended to cause females to be less receptive to remating, but different genes had effects at different timepoints. Four days after an initial mating, knockdown of *CG9897* significantly reduced the probability of remating (χ^2= 6.0; *P* = 0.01). Ten days after an initial mating, there was an interaction effect of knockdown of CG9897 and CG32834 on re-mating (χ^2= 5.2; *P* = 0.02) such that females with decreased levels of both gene products had a lower re-mating rate than would have been expected based on the effect of knockdown of each gene product individually. Knockdown of *CG32833* in males did not affect the remating rates of their mates at any time point tested.

To check for confirmation of the effects we observed in the Set 1 matings, we performed a second set of experiments (Set 2; Figure 3.5B) in which we tested for remating in females knocked-down for each of the two gene products (individually and in combination) mated to wild-type (Canton S) males. We observed the same general pattern as in Set 1 (Table 3.2 and Figure 3.5). One day after an initial mating, knockdown of *CG32834* significantly reduced the probability of remating (χ^2= 3.7; *P* = 0.05), similar to a non-significant trend observed in Set 1. Four days after an initial mating, knockdown of *CG9897* significantly reduced the probability of remating (χ^2= 6.5; *P* = 0.01). Ten days after an initial mating, there was an interaction effect of

Figure 3.5: Probability of remating by *Drosophila melanogaster* females in the presence or absence of CG32834, CG9897 and/or CG32833.

A) Remating probabilities at 1, 4 and 10 days after an initial mating for Set 1) . This set includes all eight possible combinations of gene knockdown for female genes *CG32834* and *CG9897* and male gene *CG32833.* Gene presence is indicated on the x-axis of each graph with a "+", while knockdown is indicated by a "0". **B)** Remating probabilities at 1, 4 and 10 days after an initial mating for Set 2. This set included all 4 possible combinations of gene knockdown for female genes *CG32834* and *CG9897*. All males used in this set of experiments were the wild-type Canton S stock. Error bars indicate one standard error of the mean. Statistics and sample sizes are outlined in Table 3.2.

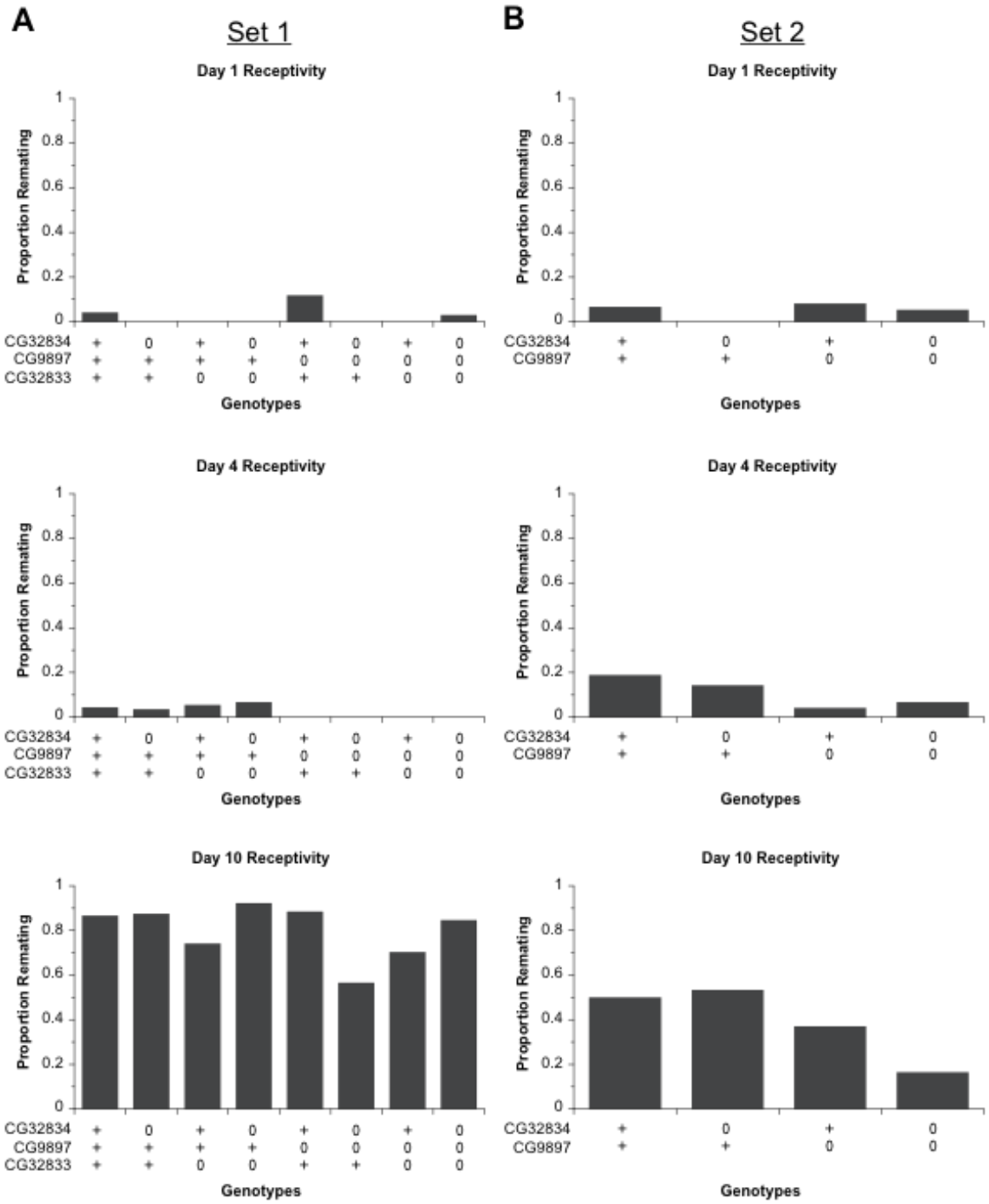

A
Set 1
Day 1 Receptivity
Proportion Remating
1
0.8
0.6
0.4
0.2
0
CG32834 + 0 + 0 + 0 + 0
CG9897 + + + + 0 0 0 0
CG32833 + + 0 0 + + 0 0
Genotypes
Day 4 Receptivity
Proportion Remating
CG32834 + 0 + 0 + 0 + 0
CG9897 + + + + 0 0 0 0
CG32833 + + 0 0 + + 0 0
Genotypes
Day 10 Receptivity
Proportion Remating
CG32834 + 0 + 0 + 0 + 0
CG9897 + + + + 0 0 0 0
CG32833 + + 0 0 + + 0 0
Genotypes
B
Set 2
Day 1 Receptivity
Proportion Remating
CG32834 + 0 + 0
CG9897 + + 0 0
Genotypes
Day 4 Receptivity
Proportion Remating
CG32834 + 0 + 0
CG9897 + + 0 0
Genotypes
Day 10 Receptivity
Proportion Remating
CG32834 + 0 + 0
CG9897 + + 0 0
Genotypes

Table 3.2: Results of analyses of remating patterns with each gene(s) knocked-down.

Gene that was knocked down	Set 1[a]			Set 2[b]		
	1 day	4 days	10 days	1 day	4 days	10 days
CG32833	n.s.[c]	n.s.	n.s.	no data[c]	no data	no data
CG9897	n.s.	χ^2 = 6.0*	n.s.	n.s.	χ^2 = 6.5**	χ^2 = 27.2***
CG32834	n.s.	n.s.	n.s.	χ^2 = 3.7*	n.s.	χ^2 = 4.7*
CG9897 and CG32834	n.s.	n.s.	χ^2 = 5.2*	n.s.	n.s.	χ^2 = 7.6**
CG9897 and CG32833	n.s.	n.s.	n.s.	no data	no data	no data
CG32834 and CG32833	n.s.	n.s.	n.s.	no data	no data	no data
CG9897, CG32834, and CG32833	n.s.	n.s.	n.s.	no data	no data	no data

[c] n.s.: not significant; no data: not tested.
* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$

Female-expressed genes are shown in pink. The male-expressed gene is shown in blue. Data were analyzed using a generalized linear model with a logit link in JMP 9. Variables were eliminated using backward iteration until only variables with P values ≤ 0.15 remained.
[a] In Set 1, control females or females knocked-down for each female gene (individually and in combination) were mated to control males or males knocked down for CG32833 for the initial mating (on Day 0) and then tested for remating with a Canton S male on the indicated day. Day 1: N = 17-37 females per treatment; Day 4: N = 15-37 females per treatment; Day 10: N = 53-61 females per treatment (two trials combined).
[b] In Set 2, control females or females knocked-down for each female gene (individually and in combination) were mated to wildtype males (Canton S) for the initial mating (on Day 0) and then tested for remating with a Canton S male on the indicated day. Day 1: N = 37-45 females per treatment (two trials combined); Day 4: N = 35-53 females per treatment (two trials combined); Day 10: N = 88-108 females per treatment (four trials combined). Contributed by L. Sirot

knockdown of *CG9897* and *CG32834* on re-mating (χ^2= 7.6; P = 0.006) such that females with decreased levels of both gene products had a lower re-mating rate than would have been expected based on the effect of knocking down each gene individually. Thus, it seems that both CG9897 and CG32834 in the female function help restore female receptivity after mating, particularly at later time points when sperm stores would be low and the advantage of remating is highest. While receipt of Sfps is known to suppress receptivity in females, the male protein CG32833 does not appear to play a role in this process.

Number of Eggs Laid

We conducted two trials to assess the effects of knocking down each gene individually and in combination on the number of eggs laid by females; results are summarized in Table 3.2 and Figure 3.6. In the first trial (Trial 1; Figure 3.6A), we compared the number of eggs laid for five days after mating between control females and females knocked-down for *CG9897* and *CG32834* individually and in combination mated to either control males or males knocked-down for *CG32833*. We found that knockdown of both *CG32834* and *CG32833* significantly reduced the number of eggs laid in the first 24 hours after mating (Figure 3.6A, top panel, and Table 3.3; CG32834: $F_{1,158}$ =27.8, P < 0.001; CG32833: $F_{1,158}$ =14.7, P < 0.001). Knockdown of both *CG9897* and *CG32833* significantly reduced the number of eggs laid over the course of days 2 to 5 after mating (Figure 3.6A, bottom panel, and Table 3.3; *CG9897*: $F_{1,158}$ =9.1, P = 0.003; *CG32833*: $F_{1,158}$ =16.7, P < 0.001). To check for confirmation of the *CG32834* and *CG9897* knockdown effects on egg-laying, we conducted a follow-up experiment (Trial 2; Figure 3.6B) in which we mated females knocked down for *CG32834* and *CG9897* (individually and in combination) to wild-type males (Canton S). Consistent across both, knockdown of *CG32834*

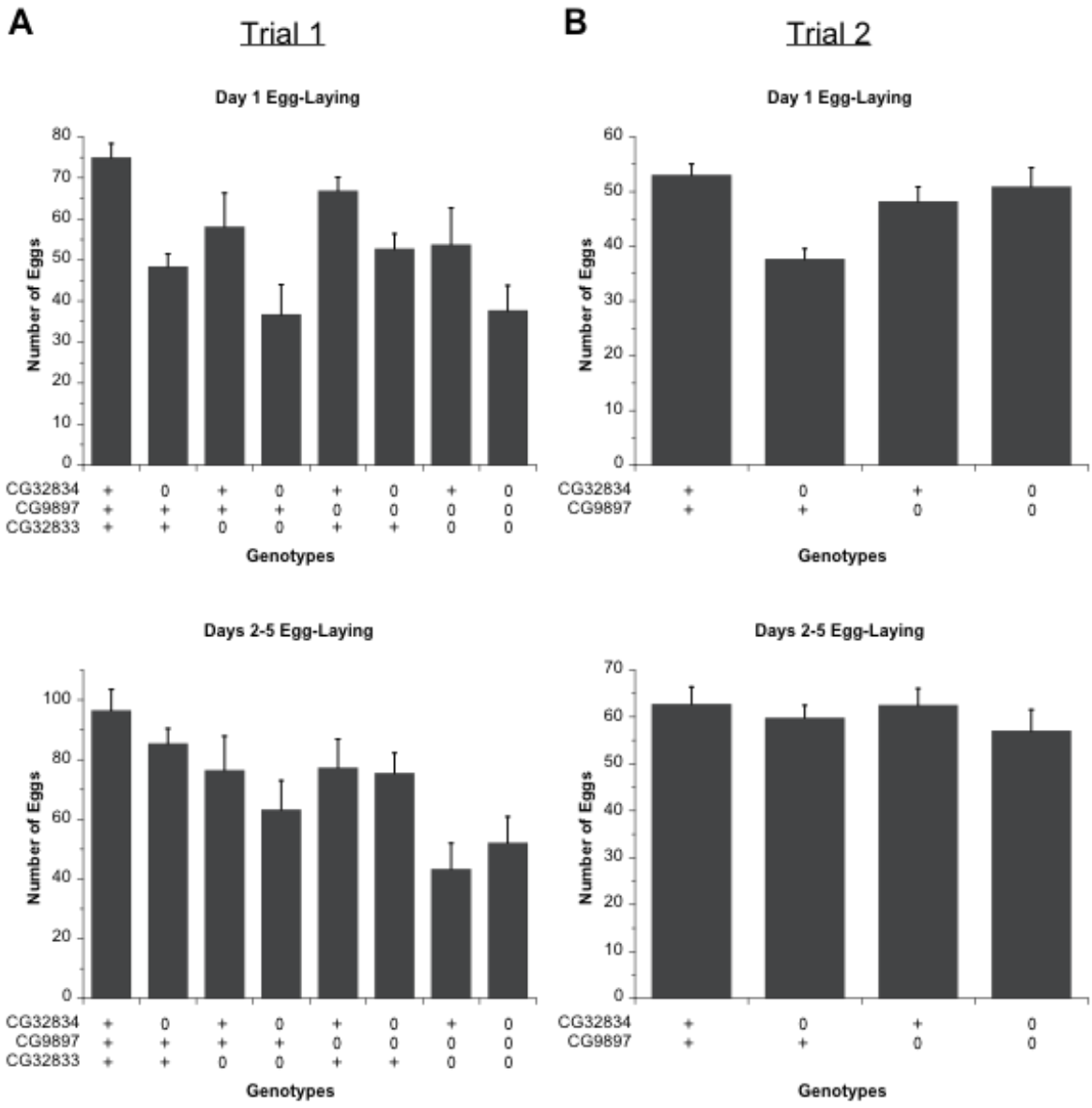


Figure 3.6: Egg-laying by *D. melanogaster* females that resulted from matings between females and males knocked down for various combinations of CG32834, CG9897 and CG32833.

A) Trial 1 used females knocked down for *CG32834*, *CG9897*, both genes, or neither gene, and males knocked down for *CG32833* or not (N=29-30 per treatment) . Top graph shows female fecundity in the first 24 hours after mating; bottom graph shows female fecundity summed over days 2-5 after mating. **B)** Trial 2 used females knocked down for *CG32834*, *CG9897*, both genes, or neither gene, mated to wild-type Canton S males (N=16-40 per treatment). Top and bottom graphs are as in (A). Gene presence is indicated on the x-axis of each graph with a "+", while knockdown is indicated by a "0". Error bars indicate one standard error of the mean.

Table 3.3: Results of analyses of number of eggs laid with each gene(s) knocked-down.

Gene that was knocked down	Trial 1		Trial 2	
	First 24 hours	Days 2-5	First 24 hours	Days 2-4
CG32833	$F_{1,158}$ =14.7***	$F_{1,158}$ =16.7***	no data	no data
CG9897	n.s	$F_{1,158}$ =9.1**	n.s	n.s
CG32834	$F_{1,158}$ =27.8***	n.s	$F_{1,131}$ =6.0*	n.s
CG9897 and CG32834	n.s	n.s	$F_{1,131}$ =12.2**	n.s
CG9897 and CG32833	n.s	n.s	no data	no data
CG32834 and CG32833	n.s	n.s	no data	no data
CG9897, CG32834, and CG32833	n.s	n.s	no data	no data

Female-expressed genes are shown in pink. The male-expressed gene is shown in blue. Data were analyzed using an ANOVA in SPSS after testing for normality with the Kolmogorv-Smirnov test..

[a] In Trial 1, control females or females knocked-down for each female gene (individually and in combination) were mated to control males or males knocked down for CG32833 (on Day 0). N = 12-30 females per treatment.
[b] In Trial 2, control females or females knocked-down for each female gene (individually and in combination) were mated to wildtype males (on Day 0). N = 15-40 females per treatment.
[c] n.s.: not significant; no data: not tested.
* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$

individually reduced the number of eggs laid in the first 24 hours after mating (Figure 3.6B, top panel, and Table 3.3; $F_{1,131}$ =6.0, P = 0.02). Together these results suggest that one of the ancestral functions of this gene family was in regulating egg-laying. Whether the ancestral gene controlled short term egg-laying or overall egg-laying is unclear, but the differences between CG32834, CG9897, and CG32833 suggest that functional divergence has occurred since the duplication events. However, these results support our hypothesis that functional Sfps can be derived from the duplication and co-option of female reproductive genes.

3.4 DISCUSSION

Gene duplication is an important evolutionary mechanism for diversifying the suite of reproductive proteins expressed within a sex (e.g., [19,21,26,33,43,56]). Instances in which reproductive gene duplication results in one paralog switching its sex-specificity of expression are less common. In one documented case, gene duplication gave rise to two cytoplasmic poly(A) RNA polymerases; one copy (WISP) retained its ancestral function in the female germline, while the other (GLD2) became expressed in the male germline [30,31,32,57]. Analogous to the case of WISP/GLD2, we have discovered a case in which a reproductive protein with ancestral female expression underwent gene duplication, with one resulting paralog becoming male-specific in expression. However, two novel features distinguish the present gene family. First, this gene family experienced a second round of gene duplication, such that all *melanogaster* subgroup species have three functional copies. Second, while the paralog *CG32833* is expressed in males, its protein product is transferred in seminal fluid to females [44]. Thus, while its pattern of expression is different, its site of action may be conserved.

Our genetic analyses of these proteins reveal that each paralog is required for normal post-mating responses in females. Females that were knocked down for one of the female genes (*CG32834*) consistently showed significantly reduced egg-laying in the first 24 hours after mating, suggesting that this protein is required for maximal early fertility of a mating pair. Furthermore, knockdown of the two female genes together consistently causes females to be significantly less likely to remate at a time point (10 days after the initial mating) at which females are often receptive to remating opportunities. Knockdown of each female gene individually also decreased the probability of remating at earlier time points (*CG32834* at one day; *CG9897* at four days). Since few *D. melanogaster* female somatic reproductive proteins have been functionally characterized [58], these experiments provide an important contribution to our understanding of the role female reproductive tract proteins play in influencing post-mating responses. Further, the male expressed duplicate CG32833 is also essential for normal egg-laying in females but does not function in regulating female receptivity. Our results suggest that in spite of millions of years of evolution and changes in sex-specificity of expression, three members of this family likely remain functional in each *melanogaster* group species.

Evolutionary History

Based on our results, this gene family appears to have a dynamic evolutionary history. Because the single copy of this gene family in *D. pseudoobscura* is expressed exclusively in the female reproductive tract, we hypothesize that the ancestral single copy of this gene was also a female-specific reproductive protein. After the divergence of *D. pseudoobscura* and the *melanogaster* species group, two gene duplication events occurred, giving rise to the three functional copies we now observe in *D. melanogaster-D. ananassae*. A fourth copy of this gene

family is annotated as *GF11311* in *D. ananassae*. Our RT-PCR and sequencing data suggest this sequence is a pseudogene. However, we found the *GF11311* transcript to be expressed specifically in male reproductive tracts (data not shown), and the corresponding amino acid sequence showed greatest identity to *GF11312*, which is expressed in both sexes. Thus, it is likely that *D. ananassae* experienced a lineage-specific duplication of its copy of the paralog that would go on to become *CG32833* in *D. melanogaster*. However, one duplicate copy has since become a pseudogene, so only one paralog remains functional today. Why and how this paralog (*GF11312*) is expressed in both sexes remains an open question, as does the exact timing of when the *CG32833* orthologs in the *melanogaster* subgroup became specifically expressed in males.

The three proteins we have described show sequence similarity to two serine-type endopeptidases, Send1 and Send2 [59]. Like CG9897 and CG32834, these proteins are also expressed specifically in the female spermathecae, and like CG9897, RNAi-induced reductions in their respective expression levels do not cause noticeable effects on female fertility or fecundity. Cloning of the Send1 and Send2 regulatory sequences showed that they have different patterns of expression: Send1 is expressed in both virgin and mated females, while Send2 is up-regulated only in mated females [59]. Send1 is also located in a gene cluster of serine-type endopeptidases with spermatheca-specific expression. These results, combined with our finding that evolutionarily related serine endopeptidases have undergone changes in sex-specific expression, suggest that the regulation of this family of reproductive serine endopeptidases may be evolutionarily labile and/or may require relatively few evolutionary steps. Cloning and functional analysis of the regulatory sequences of CG9897, CG32833 and

G32834 may lead to insights into how the expression of these genes is regulated and which *cis* regulatory elements are required for their transcription.

Our finding that a seminal fluid protein, CG32833, arose from the duplication of a female-specific reproductive protein adds to a growing body of work on the evolutionary origin of Sfps. While this mechanism of duplication of a female gene, followed by male "co-option," is intriguing, our screen for similar cases suggests it is also rare. In contrast, tandem duplication of existing Sfps is widespread in *D. melanogaster* and related species [44] and has been reported from other species (e.g., *Anopheles* mosquitoes: [43]; rodents: [60]). This dynamic process can lead to lineage-specific gene gains and losses [44,45,61]. However, other seminal proteins appear to have arisen *de novo* from non-coding regions of the genome [62,63]. Sfps in the latter class tend to be short and are often lineage-restricted, while tandem duplication of an existing protein affords the opportunity for the paralogous protein to begin with complex functional domains. It is difficult to ascertain which of these two mechanisms is more common in the origin of Sfps, but a recent genome-wide analysis in yeast suggested that *de novo* gene birth may be more common than previously thought, and possibly more common than tandem duplication in the lineages closely related to *Saccharomyces cerevisiae* [64].

Tissue-Specific Gene Expression

While tandem gene duplication is an important mechanism for generating seminal protein diversity in *Drosophila* [44], there has been little investigation of how the patterns of gene expression change or remain conserved after such duplication. In the case of CG32833, a tandem duplicate expressed in the male accessory glands but located in the genome between two female-expressed genes, a change in *cis* regulatory elements could have occurred that allowed it to

become expressed in the opposite sex from its paralogs. If both the coding sequence and the *cis* regulatory sequence were duplicated, subsequent mutations in the regulatory region of the new paralog could have resulted in altered expression. Alternatively, only the coding sequence of CG32833 could have duplicated, but this sequence could have been inserted downstream of an element that permitted a change in expression (e.g., [65]), in this case to the male accessory gland.

Another possible regulatory mechanism for this family of genes could be non-coding RNAs. The current annotation of the *D. melanogaster* genome shows two non-coding RNAs encoded in the same genomic location as the genes studied here. One is a predicted anti-sense RNA, CR42742; its sequence completely overlaps the CG9897 gene but is transcribed in the opposite direction. ModENCODE data [66] show that this transcript is expressed specifically in the male reproductive tract, where it could potentially decrease CG9897 expression levels, consistent with our observation that CG9897 is expressed in a female-biased, but not female-specific, manner. The other is an annotated microRNA, mir-4939, which overlaps the 5' end of CG32833 and is transcribed in the same direction as that gene. This microRNA was discovered in a project to catalog microRNAs in *D. melanogaster* [67]; data about its expression pattern, which could hint at a potential regulatory role, are presently unavailable.

Function

While a common outcome of gene duplication is the pseudogenization of a new paralog, three copies of these genes have been retained in diverse *Drosophila* species, and their expression patterns are largely conserved in *D. melanogaster* through *D. erecta*. Furthermore, the seminal fluid protein status of the CG32833 orthologs appears to be maintained within the

melanogaster subgroup, since the *D. yakuba* ortholog is also a seminal fluid protein [44]. Thus, there has presumably been a selective benefit to retaining the protein sequences and expression patterns of CG32834 and CG9897 (and their orthologs) in females and of CG32833 (and its orthologs) in males. Our RNAi studies of the female-expressed genes allow us to infer the reproductive benefits they may provide, at least in *D. melanogaster*. In this species, females lacking CG32834 have a reduction in fecundity (number of eggs laid) of approximately 30% on the first day after mating. Because a female's egg production is typically highest in the first 24 hours after mating, CG32834 appears to have been retained by selection for its role in boosting egg and progeny production during this critical time.

Both female-expressed genes also are required for a long-term change in female behavior. After mating once, females are typically unreceptive to remating for several days; this behavior is caused by the action of a seminal protein, sex peptide (SP), within the female [68,69]. SP binds to sperm stored in the female and is gradually released by proteolytic cleavage to prolong female post-mating responses [70]. However, by 5 days after an initial mating, the amount of SP present in the female [70] is markedly reduced, and at 10 days after a mating, females' egg production and stored sperm are at relatively low levels [71,72]. At such late time points, an additional mating would provide females with a fresh supply of sperm and seminal proteins, which would in turn increase female egg production and fertility. Remating may also benefit females indirectly through, for example, increasing the genetic diversity of her offspring or genetic compatibility with her mate. Thus, the interests of a mating pair are likely to be in conflict over whether and when the female remates. Indeed, we observed that females with reduced levels of both CG9897 and CG32834 are significantly less likely to remate 10 days after an initial mating, suggesting that the presence of these proteins promotes remating by females.

Interestingly, the two female proteins each also appear to promote remating at earlier time points (one and four days after mating). Thus, these female proteins may act individually and together to help females return to a receptive behavioral state. Notably, wild-caught female *D. melanogaster* are typically found to contain sperm from several males [73], suggesting that maintaining or regaining some level of remating receptivity after a prior mating is advantageous.

Co-Option as a Mechanism for the Evolution of "Influential" Proteins

Together, our results suggest that a seminal fluid protein that affects female post-mating responses evolved through a process of co-option in which an existing female reproductive tract protein (presumably with function in the female) was duplicated and subsequently changed expression patterns such that it was produced in the male and transferred to the female. This process of co-option may be a mechanism for the evolution of other types of "influential proteins", that is, proteins from one individual that influence the phenotype of a conspecific. Other known classes of influential proteins include those involved in various stages of sexual reproduction (e.g., courtship progression; [74]) and in maternal-fetal interactions [75]. The influences of influential proteins can be either beneficial or detrimental to the affected individual [76]. In either case, it is intriguing to consider how proteins could evolve to allow one class of individuals to influence the phenotype of another class of conspecifics. Future research should investigate whether other classes of influential proteins can also evolve through the conspecific co-option mechanisms suggested by our results for the evolution of seminal fluid proteins.

3.4 METHODS

A Targeted Search for Female Reproductive Proteins with Sex-switched Duplicates

To investigate whether female-expressed reproductive proteins might have paralogs present in male seminal fluid, we selected twenty predicted-secreted proteins expressed in sperm storage organs ([53,77,78,79]; Table 3.1). We used BLASTP to compare each protein to all other annotated proteins in *D. melanogaster*. Up to five hits per protein that showed evidence of homology (alignment score > 80; $e < 10^{-3}$, identity >30%) were checked against published data [44,53,80] for whether they were transferred in male seminal fluid or predicted to be expressed in the male accessory glands. If any candidate met either criterion, we then examined whether the proteins showed evidence of paralogy, as judged to a first approximation by reciprocal BLAST comparisons.

Identification of Orthologs

Our search identified one pair of female-derived proteins (CG9897 and CG32834) that show sequence similarity to an SFP, CG32833; the three genes encoding these proteins occur in a tightly linked cluster on chromosome 2R. Therefore, for the rest of our study, we focused on analyzing the evolutionary history and reproductive phenotypes of these three genes. We searched for orthologs of the *D. melanogaster* genes *CG9897*, *CG32833*, and *CG32834* in the 11 other *Drosophila* species that have had their genomes sequenced [55]. First, we used BLASTP to compare each gene's protein sequence from *D. melanogaster* against all predicted protein sequences from each of the other species. We detected reciprocal best BLAST proteins for one or more of the *D. melanogaster* proteins in each species from *D. simulans* to *D. pseudoobscura* and *D. persimilis*; clear orthologs were not identifiable in more distantly related species. We

found just one copy from this gene family in *D. pseudoobscura*/*persimilis*, four total copies in *D. ananassae*, and one ortholog for each *D. melanogaster* gene in each of the species from *D. simulans* to *D. erecta*.

In several instances, we found that the predicted gene models for the orthologs were incomplete or incorrect. We combined experimental approaches (PCR and sequencing) with bioinformatic alignment methods [62] to determine consistent gene structures across all species studied. For example, we found that in our lab wild-type strain of *D. melanogaster* (Canton S), the annotated intron in the CG9897 gene is instead coding DNA sequence. Furthermore, this gene region contains a 1-bp deletion relative to the FlyBase annotation. The effect of this deletion is to maintain a single open reading frame throughout the length of the coding DNA sequence, suggesting that the transcript produces a functional protein. Interestingly, while we confirmed the lack of splicing of the annotated intronic region in the strain of *D. melanogaster* used for genome sequence (*y*; *cn bw sp*), we did not observe the 1-bp deletion, suggesting that this strain carries an allele of CG9897 that has a premature stop codon that truncates the predicted protein by ~40 percent.

Phylogenetic Analysis

Once protein sequences were determined from the corrected gene structures, we constructed a phylogenetic tree from the core serine endopeptidase domain of the protein sequences we determined. This domain captured almost the entire length of the protein for the orthologs of CG9897 and CG32833, but excluded the repetitive C-terminal region of CG32834, which was difficult to align, as well as uncertain 5' ends of a few other identified copies. Sequences were aligned using CLUSTAL Omega [81] and checked by eye in the MEGA 5.05 program [82]. We

then used maximum likelihood to estimate the tree from the aligned protein sequences, using the *proml* program in PHYLIP v3.69 [83]. We visualized the tree using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). Bootstrap values, based on 100 replicates, were calculated using the PHYLIP programs *seqboot*, *proml* and *consense*.

Gene Expression

We used RT-PCR to analyze the expression of each identified ortholog in its cognate species. We used the wild-type Canton S strain for *D. melanogaster* expression and those strains from other species that were used for sequencing their genomes [55]. Using cDNA made from whole flies, we tested for sex-biased or sex-specific expression of each ortholog in its cognate species; we examined five species in the *melanogaster* subgroup (*D. melanogaster, D. simulans, D. sechellia*, *D. yakuba* and *D. erecta*), a representative of the *melanogaster* group (*D. ananassae*), and an outgroup species, *D. pseudoobscura*. We tested for tissue-specific expression in *D. ananassae*, *D. pseudoobscura*, and *D. melanogaster* as a representative of the melanogaster subgroup. (In this subgroup, gene order and whole-fly expression patterns are conserved.) To test for sex-biased or sex-specific gene expression in whole adult flies, we extracted RNA from 10 males or 10 females using the TRIzol reagent (Invitrogen). To test for tissue-specific expression, we extracted RNA from the following tissues from each sex: the gonads (10/sample), the reproductive tract without the gonads (50/sample), and the remaining carcass without the reproductive tract (10/sample). To remove genomic DNA (gDNA) remaining after RNA extraction, we incubated 1 μg of the extracted RNA with 1-2 units RQ1 DNase (Promega, Madison, WI). We then used ~0.35 μg DNase-treated RNA to synthesize cDNA, using

SmartScribe reverse transcriptase (Clontech, Mountain View, CA). The resulting cDNA was diluted 10-fold, and 1 μL was used in subsequent PCR reactions to test for gene expression.

PCR primers were designed with the Primer3 program v0.4.0 (http://frodo.wi.mit.edu/) to produce ~350 bp amplicons from cDNA. As needed, individual primers were used to sequence PCR products to confirm sequences or to identify sequences from an incorrectly annotated species. As a positive control for RT-PCR, we amplified an intron-containing region of the *RpL32* gene. To ensure that cDNA preparations were free of gDNA contamination, *RpL32* primers were designed so that the product would span an intron, thus allowing gDNA amplicons to be detected as larger fragments. In no case was gDNA contamination observed. In general, PCR products and expression patterns could be readily discerned with 30 cycles of amplification.

Functional Characterization in D. melanogaster

We tested the function of each protein from the cluster in *D. melanogaster*, since this species is the most tractable for such analysis. We used the UAS-GAL4 system to knock down genes singly or in combination. For female-expressed genes *CG9897* and *CG32834*, we mated control or knockdown females to wild-type (Canton S) males; for male-expressed *CG32833*, we mated knockdown or control males to wild-type females. We also tested the effects of mating knockdown males to knockdown females. To achieve knockdown, we used UAS RNAi-lines from the Vienna Drosophila RNAi Center [84] (VDRC; Transformant IDs: CG9897: 104987, CG32834: 46434, and CG32833: 102866). Female genes were knocked down in the spermathecae with the spermathecal-specific *Send1*-GAL4 driver (*Send1*-GAL4, *CyO*/*Gla*; [59]). We had more difficulty knocking-down the male gene (*CG32833*), but were able to eventually

accomplish knockdown. CG32833 UAS-RNAi flies were crossed to *tubulin*GAL80^{ts};*tubulin*GAL4/TM3,Sb, raised at room temperature, and shifted to the non-permissable temperature (30^{o}C) four days before eclosion. For control flies, we crossed the background stock for the RNAi lines (*y,w[1118]; P{attP,y[+],w[3`]}*) to the appropriate driver. We mated the flies in all combinations (control knock-down males to control females and to females knocked-down for each gene individually and in combination).

We used standard assays [71,72,85] to measure three phenotypes in knockdown and control flies: the probability of female remating, female fecundity (number of eggs laid) after a single mating, and egg to adult survival. We determined the probability of remating by mating females once and then testing to determine whether they would remate with a wild-type (Canton S strain) male within a 1-hour time period at 1, 4, or 10 days after the initial mating. We conducted two replicates of the assay at 10 days after the initial mating. The replicates differed in that one replicate had females that were maintained individually and transferred to new vials every day whereas the other replicate females were only transferred on day 5. We measured the number of eggs laid and egg to adult survival for eggs laid in the first 24-hours after mating and over a 5-day period after mating. Data from females that died before the end of the trial or that produced no live progeny were excluded from analyses.

Statistical Analysis

We used a generalized linear model with a logit link in JMP to test for the effect of male or female genotype (individually and in combination), trial number, and the interaction of trial number and each genotype on probability of remating. Variables were eliminated using

backward iteration until only variables with P values ≤ 0.15 remained. We used an analysis of variance (ANOVA) in SPSS to test for an effect of knocking down each gene (individually and in combination) on the number of eggs produced, after testing for normality with the Kolmogorv-Smirnov test. We used a binomial regression model in R (v. 2.15.1) to test for the effect of male or female genotype (individually and in combination) on egg to adult survival.

ACKNOWLEDGEMENTS

We thank Alex Wong for initially suggesting study of CG9897. We are grateful to M. Siegal and S. Schnakenberg for generously providing us with the Send1-GAL4 driver line and to A. Clark and the University of California San Diego *Drosophila* Species Stock Center for providing the non-*melanogaster Drosophila* species. Sequencing was performed by Cornell University's Life Science Core Laboratories Center. This work was supported by the National Institutes of Health (grant numbers: R01-HD038921 to MFW, R01-HD059060 to MFW and AG Clark, and F32GM097789 to GDF).

3.5 REFERENCES

1. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290.

2. Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154.

3. Ohno S (1970) Evolution by gene duplication. New York: Springer Verlag.

4. Roth C, Rastogi S, Arvestad L, Dittmar K, Light S, et al. (2007) Evolution after gene duplication: Models, mechanisms, sequences, systems, and organisms. Journal of Experimental Zoology Part B-Molecular and Developmental Evolution 308B.

5. Zhang JZ (2003) Evolution by gene duplication: an update. Trends in Ecology & Evolution 18.

6. Ohta T (2003) Evolution by gene duplication revisited: differentiation of regulatory elements versus proteins. Genetica 118.

7. Chan YF, Marks ME, Jones FC, Villarreal G, Jr., Shapiro MD, et al. (2010) Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. Science 327.

8. Makino T, Knowles, DG., McLysaght, A. (2010) Evolution by gene duplication; Dittmar KaL, Davis, editor. Hoboken, NJ: Wiley-Blackwell.

9. Rodin SN, Riggs AD (2003) Epigenetic silencing may aid evolution by gene duplication. Journal of Molecular Evolution 56.

10. Holland PWH, GarciaFernandez J (1996) Hox genes and chordate evolution. Developmental Biology 173.

11. Hosken DJ (2011) Gene duplication might not resolve intralocus sexual conflict. Trends in Ecology & Evolution 26.

12. Harano T, Okada K, Nakayama S, Miyatake T, Hosken DJ (2010) Intralocus Sexual Conflict Unresolved by Sex-Limited Trait Expression. Current Biology 20.

13. Stewart AD, Pischedda A, Rice WR (2010) Resolving Intralocus Sexual Conflict: Genetic Mechanisms and Time Frame. Journal of Heredity 101.

14. Gallach M, Betran E (2011) Intralocus sexual conflict resolved through gene duplication. Trends in Ecology & Evolution 26.

15. Connallon T, Clark AG (2011) The Resolution of Sexual Antagonism by Gene Duplication. Genetics 187.

16. Lewis CA, Talbot CF, Vacquier VD (1982) A PROTEIN FROM ABALONE SPERM DISSOLVES THE EGG VITELLINE LAYER BY A NON-ENZYMATIC MECHANISM. Developmental Biology 92: 227-239.

17. Swanson WJ, Vacquier VD (1995) EXTRAORDINARY DIVERGENCE AND POSITIVE DARWINIAN SELECTION IN A FUSAGENIC PROTEIN COATING THE ACROSOMAL PROCESS OF ABALONE SPERMATOZOA. Proceedings of the National Academy of Sciences of the United States of America 92: 4957-4961.

18. Vacquier VD, Swanson WJ, Lee YH (1997) Positive Darwinian selection on two homologous fertilization proteins: What is the selective pressure driving their divergence? Journal of Molecular Evolution 44: S15-S22.

19. Kresge N, Vacquier VD, Stout CD (2001) Abalone lysin: the dissolving and evolving sperm protein. Bioessays 23: 95-103.

20. Smith RC, Walter MF, Hice RH, O'Brochta DA, Atkinson PW (2007) Testis-specific expression of the beta 2 tubulin promoter of Aedes aegypti and its application as a genetic sex-separation marker. Insect Molecular Biology 16: 61-71.

21. Nielsen MG, Gadagkar SR, Gutzwiller L (2010) Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. Bmc Evolutionary Biology 10.

22. Raff EC, Hutchens JA, Hoyle HD, Nielsen MG, Turner FR (2000) Conserved axoneme symmetry altered by a component beta-tubulin. Current Biology 10: 1391-1394.

23. Nielsen MG, Caserta JM, Kidd SJ, Phillips CM (2006) Functional constraint underlies 60 million year stasis of Dipteran testis-specific beta-tubulin. Evolution & Development 8: 23-29.

24. Baker RH, Narechania A, Johns PM, Wilkinson GS (2012) Gene duplication, tissue-specific gene expression and sexual conflict in stalk-eyed flies (Diopsidae). Philosophical Transactions of the Royal Society B-Biological Sciences 367: 2357-2375.

25. Tian X, Pascal G, Monget P (2009) Evolution and functional divergence of NLRP genes in mammalian reproductive systems. Bmc Evolutionary Biology 9.

26. Mancini E, Tammaro F, Baldini F, Via A, Raimondo D, et al. (2011) Molecular evolution of a gene cluster of serine proteases expressed in the Anopheles gambiae female reproductive tract. Bmc Evolutionary Biology 11.

27. Parsch J, Meiklejohn CD, Hauschteck-Jungen E, Hunziker P, Hartl DL (2001) Molecular evolution of the ocnus and janus genes in the Drosophila melanogaster species subgroup. Molecular Biology and Evolution 18: 801-811.

28. Torgerson DG, Singh RS (2004) Rapid evolution through gene duplication and subfunctionalization of the testes-specific alpha 4 proteasome subunits in Drosophila. Genetics 168: 1421-1432.

29. Gao GJ, Cheng Y, Wesolowska N, Rong YKS (2011) Paternal imprint essential for the inheritance of telomere identity in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 108: 4932-4937.

30. Sartain CV, Cui J, Meisel RP, Wolfner MF (2011) The poly(A) polymerase GLD2 is required for spermatogenesis in Drosophila melanogaster. Development 138: 1619-1629.

31. Benoit P, Papin C, Kwak JE, Wickens M, Simonelig M (2008) PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in Drosophila. Development 135: 1969-1979.

32. Cui J, Sackton KL, Horner VL, Kumar KE, Wolfner MF (2008) Wispy, the Drosophila homolog of GLD-2, is required during oogenesis and egg activation. Genetics 178: 2017-2029.

33. Kelleher ES, Swanson WJ, Markow TA (2007) Gene duplication and adaptive evolution of digestive proteases in Drosophila arizonae female reproductive tracts. Plos Genetics 3: 1541-1549.

34. Kelleher ES, Watts TD, LaFlamme BA, Haynes PA, Markow TA (2009) Proteomic analysis of Drosophila mojavensis male accessory glands suggests novel classes of seminal fluid proteins. Insect Biochem Mol Biol 39: 366-371.

35. Kelleher ES, Pennington JE (2009) Protease Gene Duplication and Proteolytic Activity in Drosophila Female Reproductive Tracts. Molecular Biology and Evolution 26: 2125-2134.

36. Sirot LK, LaFlamme BA, Sitnik JL, Rubinstein CD, Avila FW, et al. (2009) Molecular social interactions: Drosophila melanogaster seminal fluid proteins as a case study. Adv Genet 68: 23-56.

37. Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF (2011) Insect seminal fluid proteins: identification and function. Annu Rev Entomol 56: 21-40.

38. Clauss A, Lilja H, Lundwall A (2005) The evolution of a genetic locus encoding small serine proteinase inhibitors. Biochemical and Biophysical Research Communications 333: 383-389.

39. Lin HJ, Lee CM, Luo CW, Chen YH (2005) Functional preservation of duplicated pair for RSVS III gene in the REST locus of rat 3q42. Biochemical and Biophysical Research Communications 326: 355-363.

40. Dean MD, Findlay GD, Hoopmann MR, Wu CC, MacCoss MJ, et al. (2011) Identification of ejaculated proteins in the house mouse (Mus domesticus) via isotopic labeling. Bmc Genomics 12.

41. Hurle B, Swanson W, Green ED, Sequencing NC (2007) Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. Genome Research 17: 276-286.

42. Rogers DW, Baldini F, Battaglia F, Panico M, Dell A, et al. (2009) Transglutaminase-Mediated Semen Coagulation Controls Sperm Storage in the Malaria Mosquito. Plos Biology 7.

43. Mancini E, Baldini F, Tammaro F, Calzetta M, Serrao A, et al. (2011) Molecular characterization and evolution of a gene family encoding male-specific reproductive proteins in the African malaria vector Anopheles gambiae. Bmc Evolutionary Biology 11.

44. Findlay GD, Yi XH, MacCoss MJ, Swanson WJ (2008) Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. Plos Biology 6: 1417-1426.

45. Wagstaff BJ, Begun DJ (2007) Adaptive evolution of recently duplicated accessory gland protein genes in desert drosophila. Genetics 177: 1023-1030.

46. Cirera S, Aguade M (1998) The sex-peptide gene (Acp70A) is duplicated in Drosophila subobscura. Gene 210: 247-254.

47. Ross J, Jiang H, Kanost MR, Wang Y (2003) Serine proteases and their homologs in the Drosophila melanogaster genome: an initial analysis of sequence conservation and phylogenetic relationships. Gene 304: 117-131.

48. Graham LA, Davies PL (2002) The odorant-binding proteins of Drosophila melanogaster: annotation and characterization of a divergent gene family. Gene 292: 43-55.

49. Hekmat-Scafe DS, Scafe CR, McKinney AJ, Tanouye MA (2002) Genome-wide analysis of the odorant-binding protein gene family in Drosophila melanogaster. Genome Research 12: 1357-1369.

50. Horne I, Haritos VS, Oakeshott JG (2009) Comparative and functional genomics of lipases in holometabolous insects. Insect Biochemistry and Molecular Biology 39: 547-567.

51. Sirot LK, Poulson RL, McKenna MC, Girnary H, Wolfner MF, et al. (2008) Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, Aedes

aegypti: potential tools for control of female feeding and reproduction. Insect Biochem Mol Biol 38: 176-189.

52. Sirot LK, Hardstone MC, Helinski MEH, Ribeiro JMC, Kimura M, et al. (2011) Towards a Semen Proteome of the Dengue Vector Mosquito: Protein Identification and Potential Functions. Plos Neglected Tropical Diseases 5.

53. Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39: 715-720.

54. Gelbart WM, Crosby M, Matthews B, Rindone WP, Chillemi J, et al. (1997) FlyBase: a Drosophila database. The FlyBase consortium. Nucleic Acids Res 25: 63-66.

55. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203-218.

56. Clark NL, Findlay GD, Yi XH, MacCoss MJ, Swanson WJ (2007) Duplication and selection on abalone sperm lysin in an allopatric population. Molecular Biology and Evolution 24: 2081-2090.

57. Cui J, Sackton KL, Horner VL, Kumar KE, Wolfner ME (2008) Wispy, the Drosophila Homolog of GLD-2, Is Required During Oogenesis and Egg Activation. Genetics 178: 13p.

58. Yapici N, Kim YJ, Ribeiro C, Dickson BJ (2008) A receptor that mediates the post-mating switch in Drosophila reproductive behaviour. Nature 451: 33-U31.

59. Schnakenberg SL, Matias WR, Siegal ML (2011) Sperm-Storage Defects and Live Birth in Drosophila Females Lacking Spermathecal Secretory Cells. Plos Biology 9.

60. Karn RC, Clark NL, Nguyen ED, Swanson WJ (2008) Adaptive evolution in rodent seminal vesicle secretion proteins. Mol Biol Evol 25: 2301-2310.

61. Wagstaff BJ, Begun DJ (2005) Comparative genomics of accessory gland protein genes in Drosophila melanogaster and D-pseudoobscura. Molecular Biology and Evolution 22: 818-832.

62. Findlay GD, MacCoss MJ, Swanson WJ (2009) Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. Genome Research 19: 886-896.

63. Begun DJ, Lindfors HA, Thompson ME, Holloway AK (2006) Recently evolved genes identified from Drosophila yakuba and Drosophila erecta accessory gland expressed sequence tags. Genetics 172: 1675-1681.

64. Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, et al. (2012) Proto-genes and de novo gene birth. Nature 487: 370-374.

65. Rebeiz M, Jikomes N, Kassner VA, Carroll SB (2011) Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proceedings of the National Academy of Sciences of the United States of America 108: 10036-10043.

66. Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, et al. (2010) Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science 330: 1787-1797.

67. Berezikov E (2011) Evolution of microRNA diversity and regulation in animals. Nature Reviews Genetics 12: 846-860.

68. Liu H, Kubli E (2003) Sex-peptide is the molecular basis of the sperm effect in Drosophila melanogaster. Proc Natl Acad Sci U S A 100: 9929-9933.

69. Chapman T, Bangham J, Vinti G, Seifried B, Lung O, et al. (2003) The sex peptide of Drosophila melanogaster: female post-mating responses analyzed by using RNA interference. Proc Natl Acad Sci U S A 100: 9923-9928.

70. Peng J, Chen S, Busser S, Liu HF, Honegger T, et al. (2005) Gradual release of sperm bound sex-peptide controls female postmating behavior in Drosophila. Current Biology 15: 207-213.

71. Ram KR, Wolfner MF (2007) Sustained post-mating response in Drosophila melanogaster requires multiple seminal fluid proteins. PLoS Genet 3: e238.

72. LaFlamme BA, Ram KR, Wolfner MF (2012) The Drosophila melanogaster seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. PLoS Genet 8: e1002435.

73. Imhof M, Harr B, Brem G, Schlotterer C (1998) Multiple mating in wild Drosophila melanogaster revisited by microsatellite analysis. Molecular Ecology 7: 915-917.

74. Rollmann SM, Houck LD, Feldhoff RC (1999) Proteinaceous pheromone affecting female receptivity in a terrestrial salamander. Science 285: 1907-1909.

75. Stewart F, Allen WR (1995) Comparative aspects of the evolution and function of the chorionic gonadotrophins. Reproduction in Domestic Animals 30: 231-239.

76. Wolfner MF (2009) Battle and ballet: molecular interactions between the sexes in Drosophila. J Hered 100: 399-410.

77. Allen AK, Spradling AC (2008) The Sf1-related nuclear hormone receptor Hr39 regulates Drosophila female reproductive tract development and function. Development 135: 311-321.

78. Arbeitman MN, Fleming AA, Siegal ML, Null BH, Baker BS (2004) A genomic analysis of Drosophila somatic sexual differentiation and its regulation. Development 131: 2007-2021.

79. Prokupek AM, Kachman SD, Ladunga I, Harshman LG (2009) Transcriptional profiling of the sperm storage organs of Drosophila melanogaster. Insect Molecular Biology 18: 465-475.

80. Ram KR, Wolfner MF (2007) Sustained post-mating response in Drosophila melanogaster requires multiple seminal fluid proteins. Plos Genetics 3: 2428-2438.

81. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7: 539.

82. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.

83. Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.: Distributed by the author.

84. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, et al. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448: 151-156.

85. Ram KR, Wolfner MF (2007) Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. Integrative and Comparative Biology 47: 427-445.

CHAPTER 4

EVOLUTIONARY RATE COVARIATION IDENTIFIES NEW MEMBERS OF A PROTEIN NETWORK REQUIRED FOR *DROSOPHILA* FEMALE POST-MATING RESPONSES[4]

4.1 INTRODUCTION

Sexual reproduction is a fundamental biological process by which many eukaryotic organisms transmit their genetic material to the next generation. While the end result of a successful mating is the fusion of the gametes, other molecular interactions must occur to allow this fusion. In internally fertilizing animals, males transfer to females not only sperm, but also a suite of seminal fluid proteins (Sfps) that are essential for reproductive success. Across diverse taxa, Sfps are required for: the mobilization of sperm and their storage within the female; increasing the reproductive capacity of the female; affecting the outcome of sperm competition between multiple males; and, facilitating the union of the gametes [reviewed in 1]. In insects, Sfps also alter female behaviors and physiology [2]. Effects of Sfps can be caused by interactions between specific Sfps, between Sfps and proteins on the sperm, and between Sfps and proteins native to the female reproductive tract. Thus, characterizing the functions and interactions of Sfps is important for understanding how the sexes together ensure the successful production of progeny.

[4] A version of this chapter was submitted to PLoS Genetics as Findlay, G.D., Sitnik, J.L., Aquadro, C.F., Clark, N.L., and Wolfner, M.F., Evolutionary Rate Covariation identifies new members of a protein network required for *Drosophila* female post-mating response." The covariation rate part of the paper was devised and carried out by Geoff Findlay and Nathan Clark. My contribution was the work on Intrepid (CG12558) which was originally found as part of another screen whose results are located in APPENDIX C. Figures in this chapter are labeled with their contributor. Supplementary figures that were produced by other authors and are not essential for the paper are not included.

Post-mating changes in physiology and behavior induced by Sfps have been extensively characterized in *Drosophila melanogaster* [2,3]. In response to the receipt of Sfps, females produce, ovulate and lay eggs [4,5,6]; store sperm in specialized storage organs [7,8,9,10]; show altered immune responses [11,12]; undergo changes in sleeping, feeding and excretion behavior [13,14,15,16]; and, become refractory to male courtship [17,18]. Several of these behavioral changes – egg production, sperm storage and release, and refractoriness to remating – persist in females for several days after mating and have thus been termed the long-term response [19,20,21]. The proximate cause of these changes is a short (36 amino acid) seminal protein called sex peptide (SP) [17,18]. While most Sfps are no longer detectable in females several hours after mating [22], SP persists in females for days by binding to stored sperm [19]. Gradually, the C-terminal portion of the peptide is proteolytically cleaved to release it from sperm into the female reproductive tract [19]. This C-terminal portion of SP can then signal through its receptor, sex peptide receptor (SPR), which prolongs at least some behavioral changes in the female [23,24,25,26]. Indeed, SP cleavage is required for the protein to affect female behavior for more than one day [19] and for sperm to be released efficiently from storage [27].

We have previously used RNA interference (RNAi) or gene knockout lines to test 32 Sfps for function in the SP-mediated long-term response [4,7,10,20,28,29]. These studies identified five proteins that are required for SP to function over the long term in mated females: two C-type lectins, CG1652 and CG1656; a serine protease homolog, CG9997; a cysteine-rich secretory protein, CG17575; and, a serine protease, seminase (CG10586). These proteins act in a network in which each member is required for SP to become bound to sperm [21,28]. Loss of any network protein causes an early resumption of female receptivity to remating and a decrease

in long-term fertility. Such loss also impairs the release of sperm from the seminal receptacle in the days following mating [27]. Specific members of the network act interdependently on one another. For example, males that do not produce CG9997 are unable to transfer CG1652 and CG1656 to the female, while CG1652 and CG1656 are required to slow the rate at which CG9997 is processed in the female. Thus, while SP-SPR signaling is the proximate cause of the female post-mating response, several additional Sfps are required for this signaling to persist over the long term. We refer to this set of seven proteins as the SP network.

While genomic and proteomic analyses in *D. melanogaster* have identified hundreds of proteins from sperm [30,31], seminal fluid [32,33,34,35], and the female sperm storage organs [36,37,38,39,40], we know of few examples of how these proteins interact to cause the dramatic post-mating phenotypes observed in females [21,26,28]. Biochemical approaches to identify interacting proteins are challenging due to the small amount of protein per fly, and exhaustive genetic screening of each known reproductive protein would be laborious. Here, we demonstrate a successful effort to prioritize male and female proteins for functional testing by examining covariation in their rates of evolution among species.

Evolutionary Rate Covariation (ERC) is a new metric that bioinformatically infers functional relationships between proteins based solely on their evolutionary rates across an array of species [41]. ERC operates from the hypothesis that functionally related proteins will experience correlated rate changes, because forces governing protein evolutionary rate are expected to influence entire pathways simultaneously. Evolutionary rate depends on several factors including a protein's expression level, its essentiality, and its interactions with other proteins [42,43,44,45,46,47,48,49]. Pathway-wide fluctuation in each of these factors has been

associated with correlated rate changes (i.e., ERC) between functionally related proteins [41,50,51,52,53].

In practice, an ERC value is calculated by computing the correlation between the rates of change of two proteins across all branches of a phylogeny. ERC values range from 1 to -1 for a perfect positive or negative correlation, respectively, with the genome-wide ERC distribution between all protein pairs centered at zero [41]. Functionally related pairs of proteins have been observed to have more positive ERC values in taxa as diverse as eubacteria, fungi, invertebrates and mammals [41,50,51,54,55,56,57,58]. This finding holds for proteins that share physical or genetic interactions and proteins that are found in common complexes or metabolic pathways [41,59]. Generally, a high ERC value is best interpreted as a potential functional link, which could have resulted from a common evolutionary force acting on both proteins. Accordingly, we can infer that proteins with correlated rates may be functionally related.

ERC and related methods have primarily been used to study proteins that are already known to interact functionally or physically; the use of such methods for functional prediction is only now starting to emerge [60]. We tested the utility of applying ERC prospectively by examining proteins required for *Drosophila* SP function. Because proper function of the SP network is essential for fertility, we reasoned that members of this network could have experienced shared evolutionary selective pressures over time and might thus show patterns of ERC across the phylogeny of sequenced *Drosophila* species [61]. To test this hypothesis, we created an ERC dataset specific to *Drosophila*. This analysis revealed significant levels of ERC between known members of the SP network. We then screened for new members of this network by searching for elevated ERC between known network proteins and sets of uncharacterized Sfps and female reproductive proteins. RNAi tests of 16 top candidates revealed two female and three

male proteins required for network function. Through molecular genetic analysis, we placed four of these proteins into specific positions in the SP network, and we observed that the steps in the network in which these new proteins act are largely consistent with their evolutionary correlations. Our results demonstrate that signatures of ERC can be used prospectively to predict members of a protein network, suggesting that this method may be broadly applicable for identifying novel protein interactions.

4.2 RESULTS

Proteins in the SP network show correlated evolutionary rate variation.

We first calculated Evolutionary Rate Covariation (ERC) values for all pairs of orthologous proteins (reproductive and otherwise) from 12 *Drosophila* species. Briefly, we assembled orthologous protein sequences for each gene from each species for which they were available, resulting in 11,100 multiple alignments. For each pair of alignments, we calculated the correlation coefficient between their branch-specific evolutionary rates (see Methods). The resulting ERC values ranged from -1 to 1 and reflect the degree to which evolutionary rates correlate for any particular pair of proteins. Typically, ERC values between functionally related protein pairs are elevated compared to unrelated pairs [55]. We observed this same pattern for the seven previously known members of the *Drosophila* SP network. ERC values calculated for all possible pairs of these seven proteins had a mean of 0.3115, compared to the proteome-wide mean of 0.0019. The highly significant elevation between SP network proteins (permutation p = 0.000154) suggests that ERC could be used to predict additional SP network proteins. However, since proteins that are expressed at similar levels or in similar patterns can also show correlated evolution [43], we also tested whether reproductive proteins as a class had elevated ERC values.

To do so, we examined a set of 664 proteins found in seminal fluid, sperm, or female sperm storage organs (see Methods; we refer to these proteins below as "reproductive" but note that some are also expressed in non-reproductive tissues and could thus have other functions). The mean ERC value between all reproductive proteins was 0.0326, a highly significant elevation for sets of this size (permutation $p < 0.0001$). This elevation could be driven by direct functional relationships and/or more indirect relationships such as expression patterns [41].

To control for this elevation in ERC across all reproductive proteins when evaluating correlations between individual pairs of proteins, we factored out the broad relationship between them. To do so, we recalculated ERC using only the 664 reproductive proteins to estimate the background rate of evolution, instead of all 11,100 proteins (see Methods). After this adjustment, the mean pairwise ERC between all proteins in the reproductive set fell to 0.0047. By contrast, the mean correlation between the seven known SP network proteins remained significantly elevated (mean = 0.2806; permutation $p = 0.001002$). These results suggest that while shared patterns of expression or function can cause a significant increase in ERC, a much stronger signal is shared by the specific set of proteins that act together in the SP network.

Several of the strongest pairwise correlations between known members of the SP network were found between proteins with recognized genetic interactions. For example, males that do not produce network protein CG9997 are unable to transfer CG1652 and CG1656 to females during mating [21]. These pairs of proteins show ERC values in the top 5 percent of all pairwise correlations (CG9997-CG1652: $r = 0.62$, empirical $p = 0.03$; CG9997-CG1656: $r = 0.62$, empirical $p = 0.03$; Figure 4.1). In other instances, we did not observe strong correlations between proteins that might be expected to coevolve, such as SP and SPR. However, this lack of correlation may be explained by the fact that SPR has additional ligands besides SP [62,63],

	SP	CG17575	CG1652	CG1656	seminase	CG9997	SPR
SP		0.82	-0.34	-0.09	-0.31	-0.59	0.05
CG17575	0.041		0.42	0.38	0.25	0.17	0.37
CG1652	0.744	0.118		0.47	0.53	0.62	0.26
CG1656	0.573	0.140	0.088		0.81	0.62	0.31
seminase	0.728	0.317	0.165	0.042		0.30	0.90
CG9997	0.886	0.318	0.030	0.030	0.291		-0.06
SPR	0.453	0.150	0.232	0.195	0.016	0.558	

Figure 4.1: Proteins in the SP network show a significantly elevated signature of ERC

This pairwise matrix shows ERC values (above diagonal) and their corresponding empirical *p*-values (below diagonal) between the seven known members of the SP network. Red shading indicates correlations with empirical $p < 0.05$; more intense shading indicates a stronger correlation. Figure from G. Findlay and N. Clark.

which may constrain its evolution. Nonetheless, the overall signature of correlated evolution throughout the SP network, the high proportion of positive pairwise correlations in the group (16/21), and the significant correlations between specific group members suggest that members of the SP network show significant levels of evolutionary rate covariation.

ERC reveals new candidate SP network proteins

Since we detected positive evolutionary correlations between known SP network proteins, we applied the ERC method prospectively to identify new candidate network members. For this analysis, we calculated pairwise correlations using the reproductive protein-limited data set described above, and we focused specifically on correlations between the known SP network proteins, and sets of secreted Sfps and proteins present in the female reproductive tract. To identify candidates, we queried each of five network proteins (CG1652, CG1656, CG9997, CG17575 and SP) against the sets of Sfps and female proteins, 434 in total. (SPR was not used to query the sets because it has multiple ligands [62,63], which may be expected to dampen signals of correlated evolution. Seminase was excluded because unambiguous orthologs were found in only five species, which would cause low statistical power.) We found 111 proteins (55 Sfps, 56 female proteins) that showed a significant correlation ($p < 0.05$) with at least one of the five network proteins. To further narrow this group, we focused on 21 candidates that showed a significant ($p < 0.05$) ERC with multiple SP network proteins, that showed a highly significant ($p < 0.01$) ERC with at least one network protein, and/or that belonged to a predicted functional class already known to be involved in the SP network (Table 4.1). We tested each candidate in Table 1 by using RNAi to knockdown expression of the gene in the appropriate sex; five of the 21 candidates showed no evidence of knockdown by RT-PCR and were excluded from further

Table 4.1: Candidates identified by ERC and tested for effects on 4-day remating receptivity

Expression based on data from FlyAtlas [64]. Predicted functions are from FlyBase electronic annotations. Bold indicates statistical significance for positive candidates. Abbreviations are as follows: AG = accessory gland; ST = spermatheca; FB = fat body; TG = thoracicoabdominal ganglion. Sample sizes for receptivity assay were typically ~30 per treatment; the overall range of sample sizes was 14 to 45. For examples of near-complete and partial knockdown, see Figure S4. KD: knockdown, cont: control; *p*-values are from Fisher's exact tests. (Data contributed by G. Findlay, N. Clark, and J. Sitnik)

Table 4.1: Candidates identified by ERC and tested for effects on 4-day remating receptivity

Gene Name	Predicted functional class	Expression pattern*	Significant ERC results	Amount of knockdown	4-Day Receptivity Assay
CG30433	C-type lectin	male AG	CG17575: p = 0.025 CG1652: p = 0.037 SP: p = 0.042	near-complete	KD: 21%, cont: 10% p = 0.31
CG11037	chymotrypsin-like	male AG	CG9997: p = 0.015 CG1652: p = 0.029	partial	KD: 3%, cont: 0% p = 1.00
CG11977	CRISP	male AG	CG9997: p = 0.011 CG1652: p = 0.049	near-complete	KD: 6%, cont: 8% p = 1.00
CG14034	Lipase	male AG	CG1652: p = 0.029 CG9997: p = 0.043	near-complete	KD: 30%, cont: 18% p = 0.27
CG14061 (aqrs)	serine protease homolog	male AG	CG1652: p = 0.0015 CG9997: p = 0.02 CG1656: p = 0.035	near-complete	**KD: 93%, cont: 0% p < 0.0001**
CG2975	galactosyltransferase	male AG, crop	CG17575: p = 0.003 SP: p = 0.03	complete	KD: 0%, cont: 18% p = 0.015
CG30488 (antr)	CRISP	male AG	CG9997: p = 0.009	complete	**KD: 91%, cont: 10% p < 0.0001**
CG42326	Unknown	male AG, head, eye	CG9997: p = 0.015 CG1652: p = 0.033	near-complete	KD: 6%, cont: 10% p = 0.67
CG12558 (intr)	serine protease homolog	male AG	CG9997: p = 0.007	near-complete	**KD: 79%, cont: 19% p = 0.0027**
CG42564	CRISP	male AG	CG9997: p = 0.003	near-complete	KD: 13%, cont: 6% p = 0.43
CG8420	Unknown	male AG	CG1652: p = 0.007	partial	KD: 3%, cont: 6% p = 1.00
CG13077	cytochrome b561	female ST, eye, head	CG1656: p = 0.009	near-complete	KD: 9%, cont: 27% p = 0.11
CG16713	Kunitz protease inhibitor	female ST, FB, hindgut, head, eye	CG1652: p = 0.009 CG17575: p = 0.022 CG9997: p = 0.042	none detected	n/a
CG3097	peptidase M14	female ST, hindgut, crop	CG9997: p = 0.0007 CG1652: p = 0.011	complete	KD: 3%, cont: 5% p = 1.00
CG3239 (frma)	protease/ neprilysin	female ST, FB, head, heart	CG17575: p = 0.008	partial	**KD: 70%, cont: 3% p < 0.0001**
CG4302	UDP-glucosyltransferase	female ST, MT, FB, eye, TG, head, brain	CG1656: p = 0.002 CG9997: p = 0.021	none detected	n/a
CG6910	inositol oxygenase	female ST, heart, FB	CG1656: p = 0.007 CG17575: p = 0.047	partial	KD: 3%, cont: 13% p = 0.35
CG8586	chymotrypsin-like	female ST, head, FB, eye, crop, heart	CG1656: p = 0.008 SP: p = 0.022 CG17575: p = 0.042	none detected	n/a
Mtp	phosphatidylcholine transpoter	female ST, FB, head, heart, eye, brain, TG, crop	CG1652: p = 0.041 CG9997: p = 0.048	none detected	n/a
Vkg	extracellular matrix component	female ST, FB, heart, TG, brain, head	CG17575: p = 0.007	none detected	n/a
CG5630 (hdly)	Unknown	female ST, SG, crop, tubule, hindgut, midgut	CG17575: p = 0.005	near-complete	**KD: 56%, cont: 10% p = 0.0002**

analysis. For the remaining 16 candidates, we screened for genes whose knockdown caused a significant increase in female remating receptivity four days after an initial mating. Of the 16 candidates that were at least partially knocked down by RNAi, five showed highly significant effects on 4-day remating receptivity (Table 4.1). Knockdown of the remaining 11 candidates caused no significant increase in female receptivity. This latter result could be due in some cases to insufficient knockdown or to functional redundancy with other Sfps or female proteins. Alternatively, these proteins may not function in the SP network. Of the positive candidates, three genes (*CG14061*, *CG30488* and *CG12558*) are expressed specifically in the male accessory glands [64]; at least two of them (*CG14061* and *CG30488*) encode proteins that are transferred to females as Sfps at mating [33]. The other two positive candidates, *CG3239* and *CG5630*, are each expressed in the female's spermathecae, as well as in other non-reproductive locations [64]. *CG5630* is also expressed in the female's seminal receptacle [39].

To evaluate whether each gene was required only for extended female non-receptivity, we next tested each of the five positive candidates for effects on remating receptivity at 1 day after an initial mating. As shown in Table 4.2, in no case did knockdown of a candidate gene cause an increase in short-term receptivity. Thus, rather than having general effects on female post-mating behavior, each candidate is required specifically for the long-term loss of female receptivity to remating. This phenotype is consistent with a malfunction in the SP network [20,21]. In females mated to SP network knockdown males, SP transferred at mating but not bound to sperm is sufficient for full fertility and non-receptivity 1 day after mating. However, if SP cannot bind to sperm, it is no longer detected in the reproductive tract by 4 days after mating [19]. We reasoned that if these five positive candidates affect the function of the SP network, should also affect long-term fertility, which requires the long-term storage and utilization of SP

Table 4.2: Tests of female remating receptivity 1 day after an initial mating

Gene	Results	FET *p*-value
CG14061	KD: 3/26, cont: 1/28	0.34
CG30488	KD: 0/32, cont: 4/26	0.015*
CG12558	KD: 0/14, cont: 2/15	0.48
CG3239	KD: 3/37, cont: 2/39	0.67
CG5630	KD: 1/21, cont: 1/28	1.00

*Result not in the expected direction for non-functioning SP pathway.
KD: knockdown, cont: control, FET: Fisher's exact test.
Data contributed by G. Findlay and J. Sitnik.

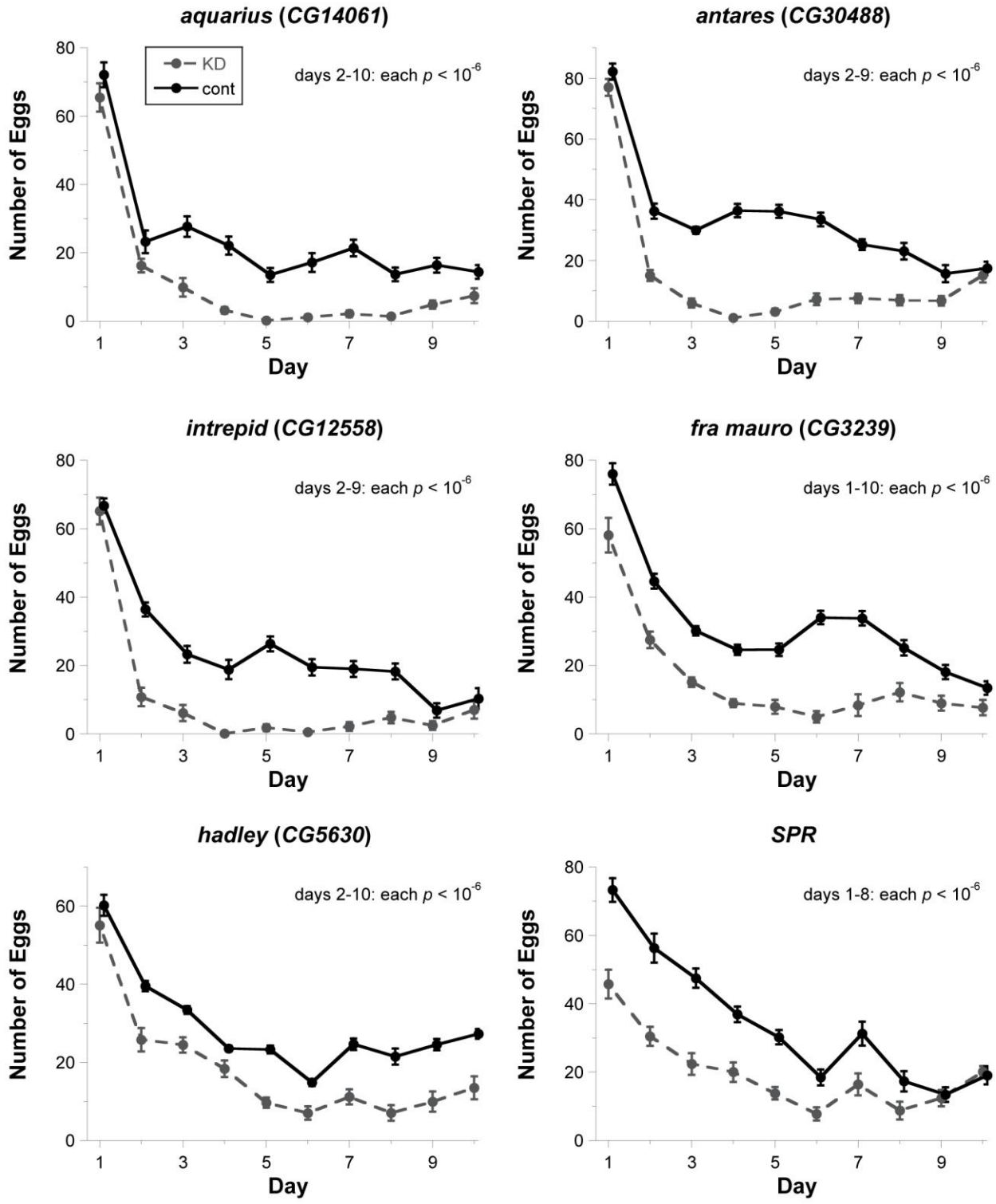
[17,18,20,26,28]. Consistent with a role in the SP network, each new protein was required for full fertility over the course of a 10-day assay (Figure 4.2). Males knocked down for *CG14061*, *CG30488* or *CG12558* induced normal levels of egg-laying and progeny production in females for the first day after mating, but these measures declined relative to controls as early as the second day after mating. Females knocked down for *CG5630* showed the same pattern of normal fertility on day 1 after mating, but reduced fertility in the following days. Females knocked down for *CG3239* had significantly reduced egg-laying and progeny production even on the first day after mating, mimicking the effects of knocking down *SPR* (Figure 4.2). These knockdown females then continued to have lower egg and progeny production throughout the assay. Furthermore, we observed that knockdown of each male gene had no significant effect on egg-hatchability, while knockdown of each female gene caused hatchability to be significantly lower (Figure 4.3). This effect was most pronounced in *CG3239* knockdown females, and much less severe in *CG5630* and *SPR* knockdown females. Effects on hatchability were unlikely to be due primarily to reduced viability of offspring inheriting both the UAS-RNAi construct and the GAL4 driver.

Thus, each of these five candidates identified by ERC is required for both the long-term loss of remating receptivity and the long-term maintenance of fertility. In our subsequent results and discussion, we adopt new names for these genes: male-expressed genes are named after lunar modules used in the Apollo space program (CG14061: *aquarius*; CG30488: *antares*; CG12558: *intrepid*), and female-expressed genes are named after sites on the moon at which Apollo missions landed (CG3239: *fra mauro*; CG5630: *hadley*).

The new male genes encode proteins predicted to belong to functional classes often found in insect and mammalian seminal fluid [33,34,65,66,67] and already represented in the SP

Figure 4.2: Fertility assays for new candidate SP network proteins identified by ERC

Each graph depicts the mean (± SE) number of eggs laid on each day of a 10-day fertility assay (knockdown: KD, dashed line; control: cont, solid line). For each male-expressed gene, knockdown or control males were mated to wild-type females. For each female-expressed gene, wild-type males were mated to knockdown or control females. Knockdown of *SPR* is shown as a comparison for *fra mauro* and *hadley* experiments. Knockdown of each gene shown had a highly significant effect (corrected $p < 10^{-6}$ in all cases) on overall fertility; results of statistical testing for fertility on each day of the assay are shown on each graph. Control data points are offset horizontally from knockdown data points to facilitate comparison, but all flies in each experiment were transferred from one vial to the next at the same time each day. Samples sizes for each treatment range from 11 to 28. One representative biological replicate (out of 2-3 for each gene) is shown. Some panels contributed by G. Findlay.


aquarius (CG14061)
KD
cont
days 2-10: each p < 10-6
antares (CG30488)
days 2-9: each p < 10-6
intrepid (CG12558)
days 2-9: each p < 10-6
fra mauro (CG3239)
days 1-10: each p < 10-6
hadley (CG5630)
days 2-10: each p < 10-6
SPR
days 1-8: each p < 10-6
Number of Eggs
Day

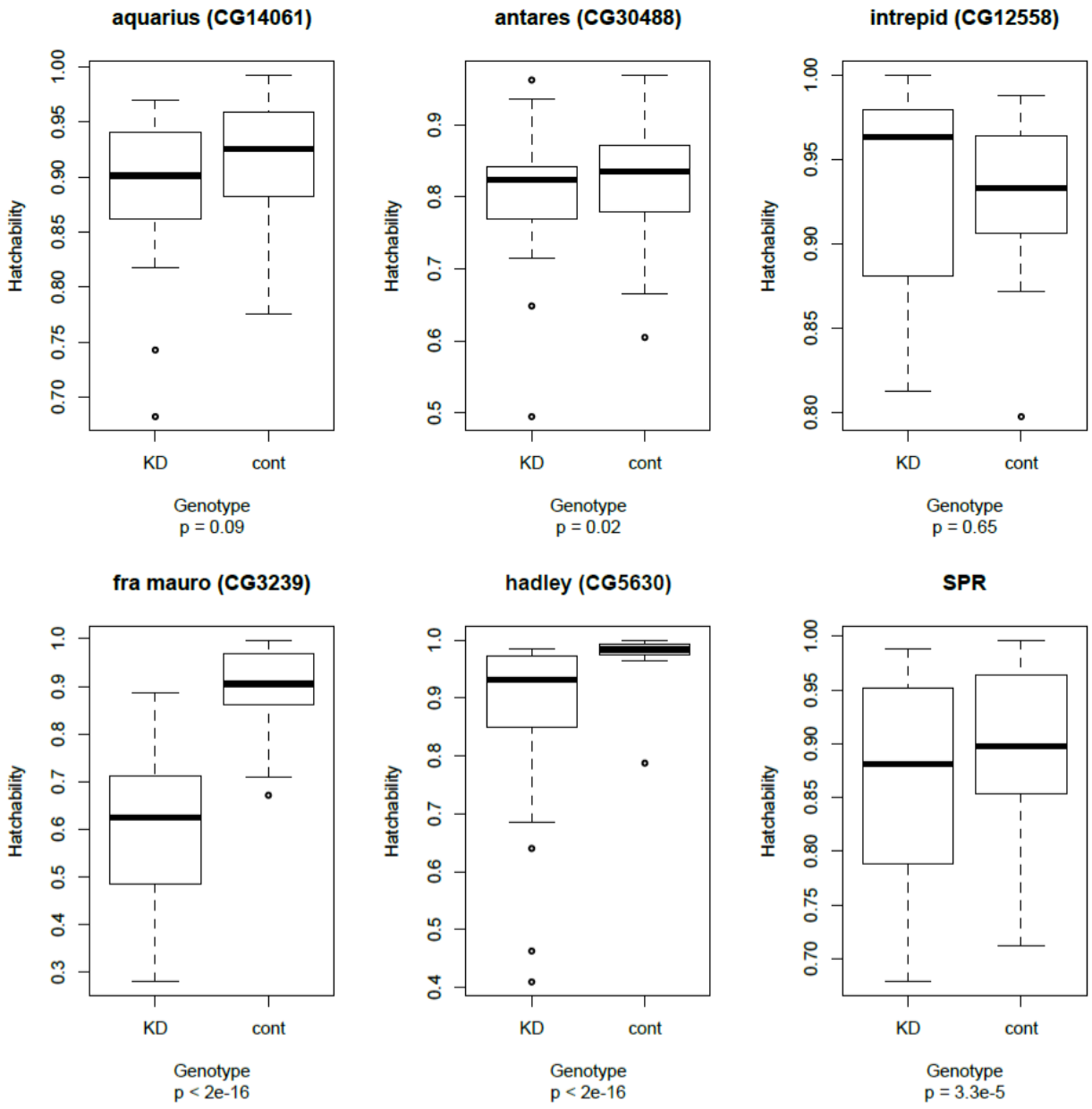


Figure 4.3: Overall rates of egg hatchability during 10-day fertility experiments

Each boxplot shows the distribution of egg hatchability rates for matings involving knockdown (KD) or control (cont) flies for each candidate gene. The thick black line represents the median rate of egg hatching across the entire 10-day assay; thin lines indicate the first and third quartiles; dots indicate outliers that lie further beyond the edge of box than 1.5x the interquartile range. *P*-values below each graph indicate results from statistical testing; after Bonferroni correction, $p < 0.0083$ are considered significant. These data come from the experiments depicted in Figure 2.

network. Like CG9997, *aquarius* and *intrepid* encode serine protease homologs [68]; like CG17575, *antares* encodes a cysteine-rich secretory protein. In females, *fra mauro* encodes a protein that contains a partial, predicted neprilysin protease domain. Neprilysins are a class of protease that preferentially cleave prohormones and neuropeptides and are important for male and female fertility in mammals [69,70,71] (Also see Chapter 2). Neither annotated isoform of fra mauro is predicted by SignalP [72] to be secreted or extracellular, raising the question of how this protein could interact with SP network proteins. Inspection of the 5' untranslated region of *fra mauro* revealed the presence of a potential alternative initiation codon, which is followed by a region predicted by SignalP to encode a functional secretion signal sequence. RT-PCR analysis on female cDNA found that a product could be amplified when a forward primer is placed in this region (data not shown), raising the possibility that an alternative isoform of the protein may be secreted and thus more accessible to other network proteins. This is similar to the state of *Nep2* (see Chapter 2), which also has a membrane bound and secreted version [73]. In addition, we found this alternative start codon and secretion signal to be conserved in 11 of 12 *Drosophila* species analyzed (the *D. willistoni* genome sequence contains a sequencing gap in this region), which provides strong evidence that this secreted protein isoform is functionally important. The hadley protein is predicted to be secreted, but its potential functional class remains unknown, as neither conserved domain searching [74] nor three-dimensional structural modeling [75] could identify a conserved protein domain.

Molecular characterization of new SP network proteins

We next sought to position these five new proteins in the SP network. To do so, we first used Western blotting to test whether SP was successfully stored over the long-term in mates of

knockdown males or in knockdown females. In wild-type matings, SP is readily detectable from dissected female seminal receptacles (SRs) 4 days after a mating. However, knockdown of any of the known SP network proteins eliminates this retention [21,28]. We observed that wild-type females mated to males knocked down for *aquarius*, *antares* or *intrepid* showed little or no SP 4 days after mating (Figure 4.4). These reduced levels of SP were not due to less SP being transferred at mating (see Figure 4). These results suggested that male proteins *aquarius*, *antares* and *intrepid* are each required for network function at a step upstream of SP binding sperm in the SR. By contrast, when wild-type males were mated to *fra mauro* or *hadley* knockdown females, normal levels of SP were observed at 4 days after mating. Thus, these two female proteins may be necessary for the utilization of SP after it becomes stored in the SR or may be required for proper SP-SPR signaling.

To further determine where the new male proteins fit into the network, we examined the production of the known SP network proteins in males knocked down for *aquarius*, *antares* or *intrepid* (Figure 4). In all cases, we observed no difference in the production of SP, CG1652, CG1656, CG9997 and CG17575 between knockdown and control males (Figure 4.5; compare lanes for knockdown and control males). We then tested whether knockdown males could transfer these proteins to females and examined their processing in female reproductive tracts. Males knocked down for *intrepid* transferred all proteins at equivalent levels to controls, and females mated to these males showed normal CG9997 processing [21] in their reproductive tracts. Males knocked down for *aquarius* or *antares* transferred normal levels of SP, CG9997 and CG17575, but much lower levels of CG1652 and CG1656 (Figure 4; compare lanes for females mated to *aquarius* or *antares* knockdown or control males). Consistent with the absence of these proteins in females after mating [21], the post-mating processing of CG9997 was also

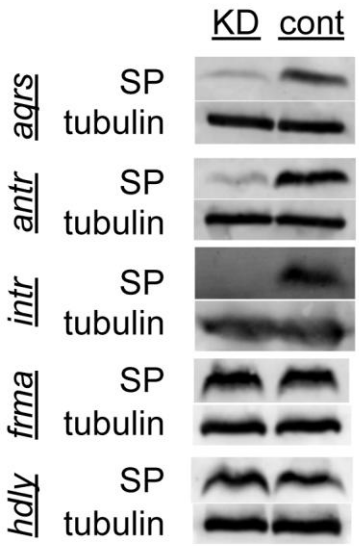


Figure 4.4: SP retention in mated females, 4 days after mating

Western blots probed with antibodies to SP or alpha-tubulin (loading control). Proteins were isolated from lower female reproductive tracts 4 days after mating. Gene names to the left of each pair of blots indicate which gene was (KD) or was not (cont) knocked down in the mating pair. Across all experiments, the number of female reproductive tract (RT) equivalents used for each condition ranged from 13 to 20; however, for any given gene, the number of RT equivalents compared between KD and control was within 2 RTs. Some panels contributed by G. Findlay.

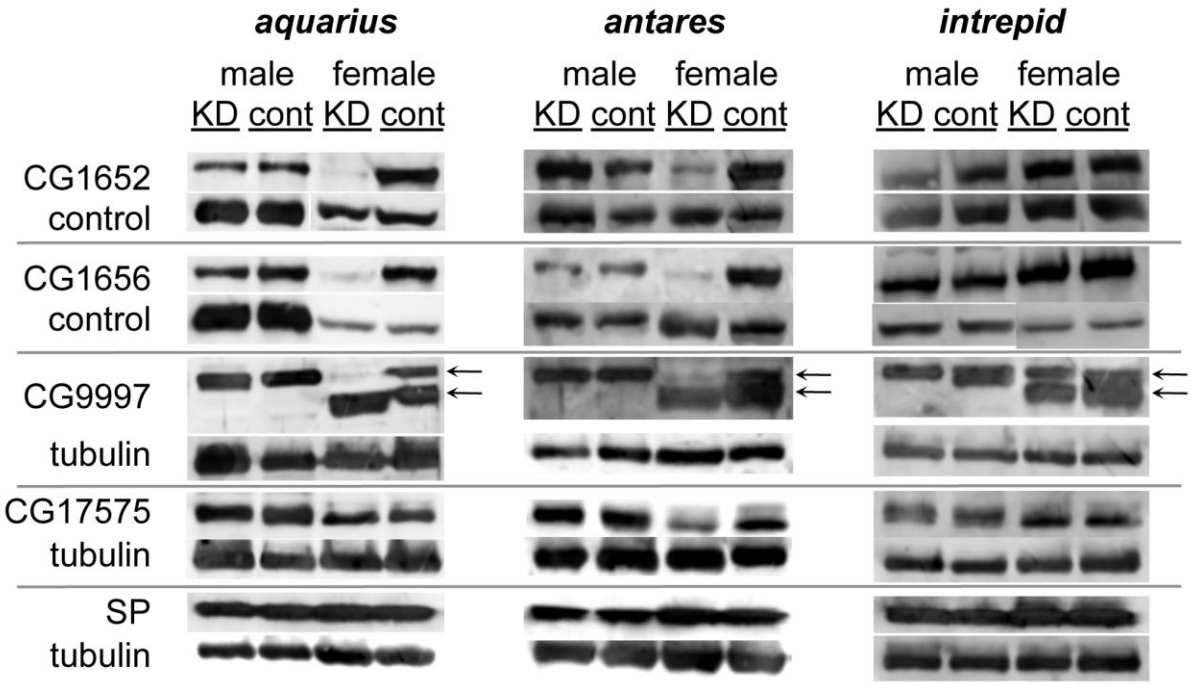


Figure 4.5: Production, transfer and processing of SP network proteins in males knocked down for *aquarius*, *antares* or *intrepid*

Western blots probed with either an antibody to an SP network protein or a loading control. Alpha-tubulin was used as the loading control for blots of CG9997, CG17575 and SP. Since CG1652 and CG1656 sometimes co-migrated with tubuiln, loading controls for these proteins were either a consistently observed cross-reactive band or tubulin. Proteins were isolated from male reproductive tracts ("male" columns) or lower female reproductive tracts dissected 1 hour after the start of mating ("female" columns). "KD" indicates males knocked down for *aqrs*, *antr* or *intr* or females mated to a knockdown male, while "cont" indicates control males or females mated to a control male. Arrows next to the blots for CG9997 indicate the ~45 (top) and ~36-kDa (bottom) forms of the protein [21]. Within each blot, the amount of RT equivalents loaded for each sex was equal. Across blots, male lanes contain 0.5-1 reproductive tract (RT) equivalents; female lanes contain 2-4 RT equivalents. Antares and Aquarius data contributed by G. Findlay.

disrupted, with mates of knockdown males showing an increased level of the 36-kDa form of CG9997 relative to the 45-kDa form of this protein. We also examined the production and transfer of seminase and observed no differences between knockdown and control flies for each gene (data not shown).

Because SP is required for the release of sperm from storage [27], we examined sperm storage and retention in the SRs of females mated to males knocked down for each of these genes (Figure 4.6). At 2 hours after mating, sperm from *antares* and *intrepid* males were present in the SR at equivalent levels to controls, while sperm from *aquarius* males were present at slightly lower levels. However, by 10 days after mating, mates of control males had largely depleted their stores of sperm in the SR, while mates of males knocked down for any of the three genes showed significantly higher numbers of sperm. Taken together with the lack of SP retention (see Figure 3), these data confirm that male proteins *aquarius*, *antares* and *intrepid* are each required for SP to become bound to sperm. Disruption of this binding, in turn, inhibits the ability of sperm to be released from the seminal receptacle. This inability to release sperm from storage likely contributes to the reduction in long-term fertility when each of these male genes is knocked down (Figure 4.2).

Taken together, our results allow us to place *aquarius*, *antares*, *fra mauro* and *hadley* into the SP network (Figure 4.7A). The male proteins *aquarius* and *antares* act at the same step of the network as CG9997, as each of these proteins is required for the transfer of CG1652 and CG1656. The female proteins *fra mauro* and *hadley* appear to act at the downstream end of the

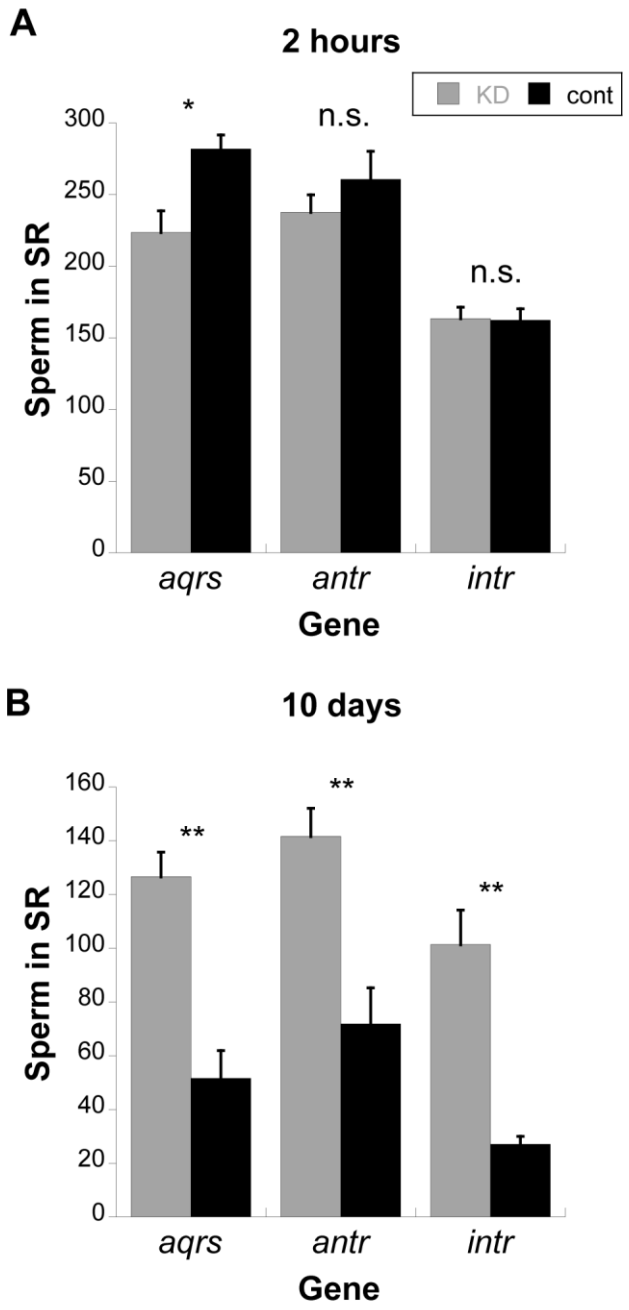
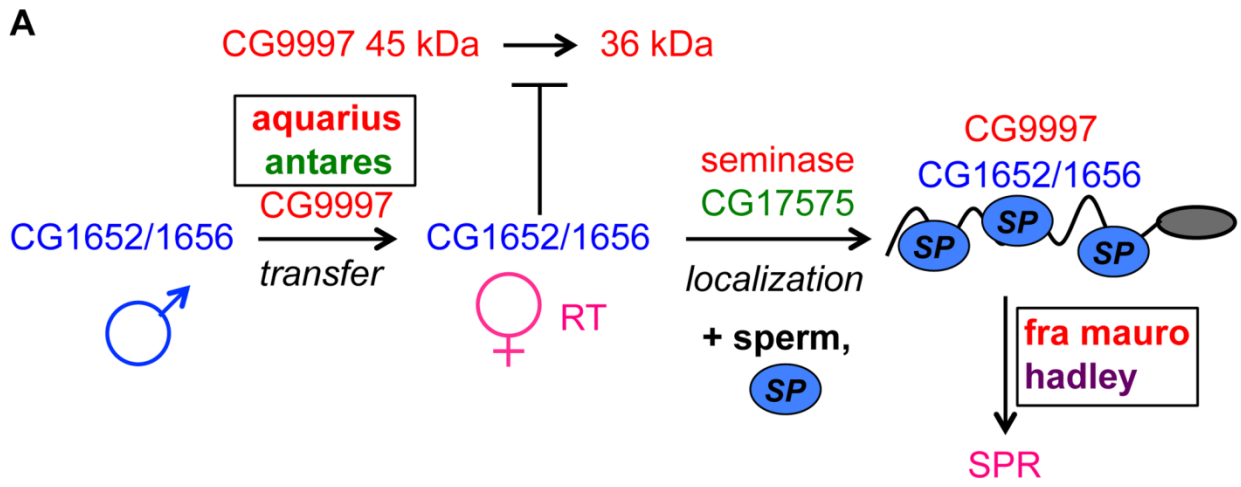


Figure 4.6: Average number of sperm stored in the seminal receptacles (SR) of wild-type females mated to knockdown or control males for new SP network proteins.

Average number of sperm in female SRs at 2 hours (A) or 10 days (B) after mating to *aqrs*, *antr* or *intr* knockdown (KD, gray) or control (cont, black) males. Each bar indicates the mean; error bars indicate 1 standard error. *, $p < 0.01$; **, $p < 0.002$; n.s. = not significant. Samples sizes for each treatment range from 11 to 18. Work contributed by G. Findlay and J. Sitnik.



B

	CG1652	CG1656	CG9997	SP	CG17575	seminase	SPR
aquarius	0.002	0.035	0.020	0.510	0.160	0.125	0.316
antares	0.153	0.225	0.009	0.159	0.444	0.847	0.652
intrepid	0.988	0.847	0.007	0.932	0.885	0.072	0.736
fra mauro	0.128	0.173	0.621	0.056	0.008	0.413	0.047
hadley	0.107	0.185	0.248	0.101	0.005	0.098	0.034

Figure 4.7: An expanded network of proteins is required for SP to bind sperm and to be utilized in mated females

A) The SP network. Colors of protein names indicate predicted protein functional classes: red = protease or protease homolog; green = cysteine-rich secretory protein (CRISP); dark blue = C-type lectin; light blue oval = SP; purple = unknown function. Boxes indicate proteins discovered by ERC; other proteins were described previously [21,28]. Intrepid acts upstream of SP-SPR signaling, but at present we cannot position it further. **B)** New members of the SP network function at steps consistent with their signals of ERC. New network proteins are shown in rows; known network proteins are shown in columns. Each cell indicates the empirical *p*-value associated with the protein's pair ERC value. *P*-values less than 0.05 are shaded in red; more intense shading indicates a stronger correlation. Work contributed by G. Findlay.

network, after SP has bound to sperm. At present, we are unable to position *intrepid* within the network, though its effect on SP retention (Figure 4.4) suggests that it acts upstream of SP-SPR signaling.

A protein's evolutionary correlations reflect its position in the SP network

When comparing the positioning of these five new proteins in the network to their patterns of ERC with the previous known seven network proteins (Figure 4.7B), we observed that the new male proteins showed their strongest correlations with the upstream players of the network. In particular, each new male protein showed a significant correlation with CG9997, which functions in the same step of the network (CG1652/CG1656 transfer) as aquarius and antares. At the downstream end of the pathway, the two new female proteins showed their strongest correlations with downstream players in the network, including SPR, which is consistent with their potential functions. Thus, the patterns of ERC observed between new and established network proteins are consistent with the steps in the network in which these new proteins are found to act.

4.3 DISCUSSION

We have used signatures of covariation in protein evolutionary rates to investigate interactions between proteins that are required to maintain post-mating responses in *Drosophila* females. We first found that, as a group, proteins known to act in the SP network [20,21,26,28] showed a significant signature of ERC. We then used ERC to screen 434 male Sfps and female reproductive tract proteins for those that correlated strongly with members of the SP network. RNAi functional testing of 16 top candidates identified five proteins that are each required for

long-lasting SP responses in females, including reducing a female's willingness to remate and boosting female egg production. The new male proteins, Aquarius, Antares and Intrepid, act in the upstream part of the network: loss of any one of these proteins prevents SP from becoming bound to sperm, which in turn prevents sperm from being released from storage. Because SP binds to sperm in females knocked down for *fra mauro* or *hadley*, these proteins may affect the ability of SP to be used in females and/or may be required for normal SP-SPR signaling. Interestingly, the strongest evolutionary correlations between these new proteins and the known members of the network occurred between pairs of proteins that appear to act in the same part of the pathway. These results verify the utility of ERC and suggest that this metric may be used prospectively to identify candidates acting in a particular part of a pathway.

ERC efficiently identifies new types of network proteins

Our results suggest that ERC successfully prioritized a large set of proteins for detailed functional testing; the observed success rate was five positive hits out of 16 candidates tested, and this rate could be higher if genetic redundancies or insufficient knockdown prevented positive results for some candidates. This rate likely represents a significant enrichment of network genes because if the same success rate were applied to the full list of 434 reproductive proteins, it would imply that there are 135 long-term mating response genes waiting to be discovered in that list alone. Although this is a formal possibility, this number seems high. Importantly, ERC allowed us to explore new functional classes of protein from the female reproductive tract. Previous studies [20,28] chose male-expressed candidates based on molecular classes that were known to function in sperm storage and fertilization. In contrast, ERC directed us to proteins that unlikely would have been selected for screening, as Fra mauro

was not annotated to be extracellular and Hadley had no predicted functional class. We can also prescribe a strategy to improve ERC analysis by retrospectively analyzing the positive candidates. Very strong correlations ($p < 0.01$) tested positive more often, so future applications of this method could focus on single, strong correlations rather than those proteins that correlate more weakly ($p < 0.05$) with multiple network members.

Possible functions for new network proteins

By expanding the SP network to include new proteins from both sexes, our results provide a more complete picture of how SP controls female post-mating responses. Until now, SPR was the only known female regulator of SP action [26], but our results show that fra mauro and hadley are also necessary for sperm-bound SP to exert its long-term effects on females. In addition to their expression in the spermathecae, both *fra mauro* and *hadley* are expressed in regions other than the female reproductive tract [64]. SPR follows the same pattern: it is expressed in several reproductive regions [26], including the spermathecae, and elsewhere in the adult female. However, only six SPR-expressing neurons in the reproductive tract are required for the SP response [23,24,25]. It is also interesting to compare the fertility phenotypes for *fra mauro*, *hadley* and *SPR* knockdown females (Figure 4.2). Knockdown of *fra mauro* or *SPR* causes both a long-term fertility deficit and an immediate reduction in egg-laying in the first 24 hours after mating. In contrast, *hadley* knockdown females show normal fertility on day 1, but then have reduced fertility over the following days. One possible model to explain these results could be that Fra mauro is necessary to facilitate SP-SPR signaling, while Hadley is necessary for the efficient release of SP from stored sperm. SP-SPR signaling is required for full fertility at all time points after mating [26] (Figure 2), but impaired release of SP from sperm affects

fertility only after day 1 [19]. Another possibility is that Fra mauro is required to coordinate temporally the release of sperm from storage when eggs are ovulated and ready to be fertilized. While knockdown of *fra mauro*, *hadley* and *SPR* each caused a reduction in egg hatchability, the magnitude of this effect was by far the greatest for *fra mauro* (Figure 4.3). Thus, in addition to laying significantly fewer eggs than controls (Figure 4.2), *fra mauro* knockdown females also experience far lower egg-to-adult viability.

Two observations suggest that interactions between SP network proteins may begin in the male. First, CG9997, Aquarius and Antares are each required for lectins CG1652 and CG1656 to be transferred efficiently to females [21] (Figure 4.5). It is possible that one or more of the former proteins may bind to either lectin protein as Sfps transit the male reproductive tract during mating. Such binding could protect the lectins from proteolysis or modification. For instance, *CG9997* and *aquarius* both encode serine protease homologs predicted to have inactivating mutations in their active sites [68]. It has been speculated that such inactive proteases could act as competitive inhibitors of proteolytic processing by binding to processing targets, rendering them less accessible to the numerous active protease in the seminal fluid [76]. Second, it is presently unclear whether intrepid is transferred at mating, as previous proteomic experiments have not detected this protein in mated females [33]. While intrepid may be transferred but poorly detectable in mated females (e.g., due to low abundance or rapid degradation), it may, alternatively, act in males to modify or activate another network protein(s). Processing of Sfps within males is observed in other cases. For example, the *Drosophila* seminal metalloprotease CG11864 is processed in the male reproductive tract during transfer to females [28,77], and this processing is required for CG11864 to mediate the processing of additional Sfps in the female reproductive tract. In nematodes, interactions between a protease, TRY-5, and a

protease inhibitor, SWM-1, regulate the activation of sperm during transit through the male reproductive tract [78,79,80]. Thus, it will be of interest to determine whether any members of the SP network are the agents or targets of processing within the male reproductive tract.

Evolution of the SP network

Our results, combined with previous work [20,21,26,81], suggest that at least 12 proteins participate in the SP-mediated post-mating response in female *Drosophila melanogaster*. How did this complex network arise, and how have its members evolved? Orthologs of the sex peptide receptor (SPR) are found in diverse insect taxa, including mosquitoes, silkworms and moths, and these receptors are responsive to stimulation by *D. melanogaster* SP [26,82]. However, SP has not been identified outside of Diptera; a putative SP ortholog was identified by bioinformatics in *Anopheles* [83], but the short length of SP makes it difficult to detect orthologs in other species, including some drosophilids. Furthermore, the female post-mating responses of insects with SPR orthologs often differ substantially from those of the *melanogaster* group of *Drosophila*. For example, *D. mojavensis* females re-mate more readily than *D. melanogaster* females [84], and while *An. gambiae* females become unreceptive to further courtship after a single mating, this behavioral change does not require the transfer of sperm [85].

Within the genus *Drosophila*, other members of the network show different levels of evolutionary conservation. We identified orthologs of CG1652, CG1656, CG9997 and CG17575 in 11 of 12 sequenced *Drosophila* species (all but the most distant species, *D. grimshawi*). Most of the new proteins we identified share this broad distribution throughout the genus. Hadley and Fra mauro are found in all 12 species, but appear not to have orthologs in sequenced mosquito species (data not shown). *Aquarius* and *antares* show the same species distribution as CG1652,

CG1656, CG9997 and CG17575. Only *intrepid* and *seminase* appear to have evolved more recently; orthologs of the former were found in 9 of 12 species (all but *D. virilis, mojavensis* and *grimshawi*), while seminase orthologs were detected only in *D. melanogaster-D. ananassae*. Taken together, these varying degrees of evolutionary conservation suggest that the SP network, as it presently functions in *D. melanogaster*, may have evolved in pieces over time.

Reproductive proteins of many species have evolved under positive selection [86,87,88]. One proposed explanation for this pattern suggests that males and females may experience sexual conflict over some aspect of reproduction (e.g., the rate of female remating). Substantial evidence suggests that sexual conflict occurs in *D. melanogaster* [89,90,91] and is mediated by SP [92]. At the molecular level, the result of sexual conflict could be continual coevolution between male and female protein sequences. Population genetic studies have detected evidence of recent selective sweeps on SP [93] and CG9997 [94], but most other members of the network appear well conserved [33]. One possible explanation centers on the observation that SPR is sensitive to multiple ligands [26,62,63], which may constrain its ability to coevolve with SP and thus reduce the requirement for constant coevolution. It will also be instructive to examine the molecular evolution of all network members across the *Drosophila* phylogeny and to determine whether any have experienced bursts of positive selection on the same phylogenetic lineages, as might be predicted for proteins showing patterns of ERC [50].

Finally, we observed several cases in which two SP network proteins are encoded by adjacent genes in the *D. melanogaster* genome (Table S2). This pattern was previously observed for the SP network lectins, *CG1652* and *CG1656*, which are believed to have arisen from an ancient gene duplication event [33,34]. We found that three additional pairs of network genes (*CG9997* and *aquarius*, *CG17575* and *antares*, and *fra mauro* and *SPR*) are also located in

tandem with one another; however, in no case do we observe unambiguous evidence for the cluster arising by tandem gene duplication. It is possible that such genomic clustering enables genes that function in a common pathway to be co-regulated [95].

Conclusions

We have shown that signatures of evolutionary rate covariation can be used prospectively to identify new members of a protein network. In the context of the *Drosophila* SP pathway, this genomic approach allowed us to efficiently screen hundreds of known reproductive proteins so as to prioritize candidates for functional analysis, thereby identifying new long-term mating response proteins from both males and females. Interestingly, male and female proteins appear to participate in distinct sections of the SP network, and this separation was reflected in their signatures of correlated evolution. We believe that the ERC approach will be broadly applicable to identifying new members of other protein networks in any taxa for which comparative genomic data are available.

Acknowledgements

We thank Norene Buehner for assistance with experiments and Brooke LaFlamme for assistance with statistical analysis. We are grateful to Laura Harrington for letting us use her microscope for sperm counting. We thank the Vienna *Drosophila* RNAi Center, the Bloomington Stock Center, and the TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing transgenic RNAi fly stocks and/or plasmid vectors used in this study. Comments from PRV Satyaki and the Wolfner Lab improved the manuscript. This research was supported by NIH grants to MFW (R01HD038921), MFW + AG Clark (R01HD059060), and

CFA (R01GM64313), NIH National Research Service Awards to GDF (F32GM097789) and NLC (F32GM084592), and a Cornell Center for Comparative and Population Genomics fellowship to GDF. For part of this work JLS was supported on National Institutes of Health training grant T32-GM07617

4.4 METHODS

Reproductive proteins data sets

We used a combination of published proteomic and transcriptomic data sets and genome-wide expression data to create three sets of reproductive genes used in the analysis: seminal fluid proteins (Sfps), female reproductive tract proteins, and sperm proteins. The first set consisted of 208 genes encoding Sfps that had been identified by mass spectrometry in the reproductive tracts of mated females [32,33] or predicted secreted proteins from the male accessory gland [34]. The second set included 226 genes expressed in the female sperm storage organs. This set included the *D. melanogaster* orthologs of EST sequences identified from the spermathecae of *D. simulans* [36,38] and EST sequences identified from the seminal receptacle of *D. melanogaster* [39]. We removed from these sets annotated housekeeping genes (e.g., ribosomal and mitochondrial proteins) since they were unlikely to interact with proteins in the SP network. Because EST sequencing may not sample all relevant genes, we then supplemented these genes with genes identified in FlyAtlas [64] to be predominantly expressed in the spermathecae (the only female sperm storage organ for which genomewide expression data are available). The third set included 322 genes that encode proteins in the *D. melanogaster* sperm proteome [30,31] and that were found in FlyAtlas to be predominantly expressed in the testis. This filtering was performed to enrich for proteins likely to function specifically in reproduction, since proteins

involved in additional biological processes may interact with several partners and thus show dampened signals of ERC. While we used all three sets of genes (756 genes in total) for optimizing the ERC method (see below), we focused our further functional tests on ERC candidates identified from the seminal fluid and sperm storage organ gene sets (434 in total).

Alignment of orthologous protein coding sequences from 12 species

We identified orthologous genes from 12 *Drosophila* species using a combination of high-throughput and manual searching. Protein amino acid sequences were produced by the *Drosophila* 12 Genomes project and downloaded from FlyBase (http://flybase.org) [61]. The species were: *Drosophila melanogaster, sechellia, simulans, yakuba, erecta, ananassae, pseudoobscura, persimilis, willistoni, grimshawi, virilis,* and *mojavensis.* Orthologs were identified using InParanoid, and the resulting groups were aligned by MUSCLE [96,97]. Many alignments were missing species either due to evolutionary loss or missed gene annotation. To increase the number of species and thereby improve our power, we manually searched for unannotated genes in the 11 non-*melanogaster* species using a combination of tBLASTn and BLAT. This effort added 81 previously unannotated sequences to a total of 31 alignments.

Genome-wide Evolutionary Rate Covariation (ERC) analysis across 12* Drosophila *species

To perform ERC analysis, we first calculated the amount of amino acid divergence for each branch in the species tree for each of the 11,100 orthologous protein alignments produced above; this was done using 'aaml' of the PAML package [98]. Next, raw branch lengths were transformed into rates of evolution relative to the expected branch length. This projection operation, introduced by Sato et al. [58], removes the inherent correlation of all proteins due to

the underlying species tree and improves the power of ERC to resolve functionally related protein pairs from unrelated pairs [55,58]. Finally, we used these corrected branch-specific rates to calculate the correlations for all pairs of proteins, resulting in a proteome-by-proteome matrix of correlation coefficients, termed the ERC matrix. To limit the effect of outlier points, we limited all rates to 2 standard deviations from the mean.

In spite of our efforts (above) to improve species coverage, most alignments were missing at least one species. We set a minimum species threshold at 5, so species representation ranged from 5 to 12. This heterogeneity required us to create a flexible system to compare ERC results between different sets of species. A table of relative rates (projection operation, above) was produced for each unique set of species shared between protein pairs, resulting in 1,815 projections. Importantly, the distribution of ERC values varied depending on the particular set of species employed. For example, the variance of ERC values is consistently larger for smaller numbers of species (Figure S3). To correct for these effects we converted every observed ERC value in to an empirical *p*-value based on the observed distribution of ERC values for that particular set of species. The comparison of *p*-values allowed us to compare ERC results across all protein pairs. Hence, we report all ERC results as *p*-values ranging from 0 to 1, where a lower value indicates stronger evidence for rate correlation.

Significance testing for sets of proteins was performed by comparing the mean ERC value between all pairs in the set to permutations of randomly chosen ERC values from the same species-matched projections. This exact matching of projections ensured that the random permutations were taken from the exact same distributions as the observed values.

The “reproductive protein only” analysis was performed as above, except that analysis was limited to the 664 Sfps, female proteins, and sperm proteins described above with at least 5

species. Significance testing for single pairs and for sets of proteins was performed as above, through empirical *p*-values. Calculations of pairwise correlations between pairs of known network proteins and between known network proteins and members of the sets of Sfps and female proteins were performed using this reproductive protein set.

RNA interference (RNAi)

To knock down expression of candidate genes, we used a variety of RNAi lines and drivers. Most lines were second-generation (KK) RNAi lines provided by the Vienna *Drosophila* RNAi Center (www.vdrc.at) [99]; several others were either provided by the Transgenic RNAi Project (TRiP; Harvard University) [100] or constructed in house using the pVALIUM20 vector [101,102] provided by the TRiP. When possible, we used the *tubulin*-GAL4 driver to knockdown genes ubiquitously, but in some cases knockdown with this driver caused lethality. When ubiquitous knockdown of a male-expressed Sfp gene caused lethality, we first attempted to use the *prd*-GAL4 driver [103] to knockdown expression in the accessory glands. However, we observed phenotypes consistent with SP network malfunction when this driver was crossed to a control background strain that does not induce RNAi. Thus, we instead used the *ovulin*-GAL4 driver [17] to knock down male Sfp genes. To knockdown female genes expressed in the spermathecae, we used the *Send1*-GAL4 driver [104], sometimes in combination with a UAS-Dicer2 sequence to enhance RNA interference. The RNAi line numbers, specific crosses and genetic controls used are given in Table S1. All flies were reared on a 12 hr/12 hr light-dark cycle. Most crosses were performed at room temperature (22°C ± 1°); some were instead performed at 25° to attempt to induce greater knockdown.

We determined the degree of knockdown by using RT-PCR [20,28] to measure the expression level of each RNAi-targeted gene in knockdown flies and their respective controls, using amplification of the *RpL32* transcript as a positive control (see Supporting Methods for further details). For *tubulin*-GAL4 knockdown, we analyzed RNA isolated from whole flies; for tissue-specific knockdown, we analyzed RNA isolated from dissected reproductive tracts. We qualitatively scored the degree of knockdown as "complete/near complete," "partial," or "no detectable knockdown", and we chose for functional analyses only those genes (16 of 21 tested) that showed at least partial knockdown. Figure S4 shows knockdown levels for all positive candidates.

Screens for reproductive phenotypes

For several days after an initial mating, females are reluctant to remate in a one-hour, single-pair test, but only if the SP network is functioning properly [19,20]. Thus, we initially screened each candidate gene for its effects on a female's willingness to remate within 1 hour, 4 days after an initial mating, using previously described methods [20]. Positive candidates were then evaluated by the same assay for remating receptivity at 1 day after mating, and for fertility, fecundity and egg hatchability over 10 days after an initial mating. These assays were performed according to previously described methods, with minor modifications. For more detail, see Supporting Material.

Western blotting

To examine the production, transfer and processing of known SP network proteins in flies knocked down for a newly identified candidate, we performed Western blot experiments using

available antibodies to SP, CG1652, CG1656, CG9997 and CG17575 as previously described [21]. For each positive candidate, we first tested whether SP was retained on sperm over the long term by dissecting 13-20 lower female reproductive tracts for each treatment at 4 days ASM. While the number of female reproductive tracts per lane across experiments varied within this range, pairs of samples being compared never differed by more than 2 tracts. Extracted proteins were run on 15% acrylamide gels, transferred to membranes, and then probed for SP and alpha-tubulin (as a loading control) as previously described.

For candidates that caused a reduction of SP levels in females at 4 days ASM, we then evaluated the production, processing and transfer of the known network proteins by testing for their presence in male reproductive tracts and in mated females at 1 hr ASM. Proteins were separated on 10.6% acrylamide gels and then transferred and probed for as described previously. Approximately 0.5-1 male reproductive tract equivalents and 2-4 lower female reproductive tract equivalents were loaded in each lane. While the number of female reproductive tract equivalents per lane varied between blots for different SP network proteins, comparisons between knockdown and control flies for any given protein were performed with an equal number of reproductive tracts in each lane. As a loading control for each blot, we primarily used alpha-tubulin. In cases where CG1652 and CG1656 co-migrated with alpha-tubulin, we also examined a consistently observed cross-reactive band.

4.5 REFERENCES

1. Poiani A (2006) Complexity of seminal fluid: a review. Behavioral Ecology and Sociobiology 60: 289-310.

2. Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF (2011) Insect seminal fluid proteins: Identification and function. In: Berenbaum MR, Carde RT, Robinson GE, editors. Annual Review of Entomology, Vol 56. pp. 21-40.

3. Chapman T, Davies SJ (2004) Functions and analysis of the seminal fluid proteins of male Drosophila melanogaster fruit flies. Peptides 25: 1477-1490.

4. Herndon LA, Wolfner MF (1995) A *Drosophila* seminal fluid protein, Acp26Aa, stimulates egg-laying in females for 1 day after mating. Proceedings of the National Academy of Sciences of the United States of America 92: 10114-10118.

5. Soller M, Bownes M, Kubli E (1997) Mating and sex peptide stimulate the accumulation of yolk in oocytes of *Drosophila melanogaster*. European Journal of Biochemistry 243: 732-738.

6. Soller M, Bownes M, Kubli E (1999) Control of oocyte maturation in sexually mature *Drosophila* females. Developmental Biology 208: 337-351.

7. Avila FW, Wolfner MF (2009) Acp36DE is required for uterine conformational changes in mated *Drosophila* females. Proceedings of the National Academy of Sciences of the United States of America 106: 15796-15800.

8. Bloch Qazi MC, Wolfner MF (2003) An early role for the *Drosophila melanogaster* male seminal protein Acp36DE in female sperm storage. Journal of Experimental Biology 206: 3521-3528.

9. Neubaum DM, Wolfner MF (1999) Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153: 845-857.

10. Wong A, Albright SN, Giebel JD, Ravi Ram K, Ji S, et al. (2008) A role for Acp29AB, a predicted seminal fluid lectin, in female sperm storage in *Drosophila melanogaster*. Genetics 180: 921-931.

11. Peng J, Zipperlen P, Kubli E (2005) *Drosophila* sex peptide stimulates female innate immune system after mating via the Toll and Imd pathways. Current Biology 15: 1690-1694.

12. Short SM, Lazzaro BP (2010) Female and male genetic contributions to post-mating immune defence in female *Drosophila melanogaster*. Proceedings of the Royal Society B-Biological Sciences 277: 3649-3657.

13. Carvalho GB, Kapahi P, Anderson DJ, Benzer S (2006) Allocrine modulation of feeding behavior by the sex peptide of *Drosophila*. Current Biology 16: 692-696.

14. Cognigni P, Bailey AP, Miguel-Aliaga I (2011) Enteric neurons and systemic signals couple nutritional and reproductive status with intestinal homeostasis. Cell Metabolism 13: 92-104.

15. Isaac RE, Li C, Leedale AE, Shirras AD (2010) *Drosophila* male sex peptide inhibits siesta sleep and promotes locomotor activity in the post-mated female. Proceedings of the Royal Society B-Biological Sciences 277: 65-70.

16. Apger-McGlaughon J, Wolfner MF (2013) Post-mating change in excretion by mated Drosophila melanogaster females is a long-term response that depends on sex peptide and sperm. J Insect Physiol.

17. Chapman T, Bangham J, Vinti G, Seifried B, Lung O, et al. (2003) The sex peptide of *Drosophila melanogaster*: Female post-mating responses analyzed by using RNA interference. Proceedings of the National Academy of Sciences of the United States of America 100: 9923-9928.

18. Liu HF, Kubli E (2003) Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America 100: 9929-9933.

19. Peng J, Chen S, Busser S, Liu HF, Honegger T, et al. (2005) Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. Current Biology 15: 207-213.

20. Ravi Ram K, Wolfner MF (2007) Sustained post-mating response in *D. melanogaster* requires multiple seminal fluid proteins. Plos Genetics 3: e238.

21. Ravi Ram K, Wolfner MF (2009) A network of interactions among seminal proteins underlies the long-term postmating response in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America 106: 15384-15389.

22. Ravi Ram K, Ji S, Wolfner MF (2005) Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaster*. Insect Biochemistry and Molecular Biology 35: 1059-1071.

23. Hasemeyer M, Yapici N, Heberlein U, Dickson BJ (2009) Sensory neurons in the *Drosophila* genital tract regulate female reproductive behavior. Neuron 61: 511-518.

24. Rezaval C, Pavlou HJ, Dornan AJ, Chan Y-B, Kravitz EA, et al. (2012) Neural circuitry underlying *Drosophila* female postmating behavioral responses. Current Biology 22: 1155-1165.

25. Yang CH, Rumpf S, Xiang Y, Gordon MD, Song W, et al. (2009) Control of the postmating behavioral switch in *Drosophila* females by internal sensory neurons. Neuron 61: 519-526.

26. Yapici N, Kim YJ, Ribeiro C, Dickson BJ (2008) A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. Nature 451: 33-U31.

27. Avila FW, Ram KR, Qazi MCB, Wolfner MF (2010) Sex peptide is required for the efficient release of stored sperm in mated *Drosophila* females. Genetics 186: 595-600.

28. LaFlamme BA, Ram KR, Wolfner MF (2012) The *Drosophila melanogaster* seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. Plos Genetics 8.

29. Mueller JL, Linklater JR, Ravi Ram K, Chapman T, Wolfner MF (2008) Targeted gene deletion and phenotypic analysis of the *Drosophila melanogaster* seminal fluid protease inhibitor Acp62F. Genetics 178: 1605-1614.

30. Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, et al. (2006) Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. Nature Genetics 38: 1440-1445.

31. Wasbrough ER, Dorus S, Hester S, Howard-Murkin J, Lilley K, et al. (2010) The *Drosophila melanogaster* sperm proteome-II (DmSP-II). Journal of Proteomics 73: 2171-2185.

32. Findlay GD, MacCoss MJ, Swanson WJ (2009) Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. Genome Research 19: 886-896.

33. Findlay GD, Yi X, MacCoss MJ, Swanson WJ (2008) Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. Plos Biology 6: 1417-1426.

34. Ravi Ram K, Wolfner MF (2007) Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduuction. Integrative and Comparative Biology 47: 427-445.

35. Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America 98: 7375-7379.

36. Allen AK, Spradling AC (2008) The Sf1-related nuclear hormone receptor Hr39 regulates *Drosophila* female reproductive tract development and function. Development 135: 311-321.

37. Arbeitman MN, Fleming AA, Siegal ML, Null BH, Baker BS (2004) A genomic analysis of *Drosophila* somatic sexual differentiation and its regulation. Development 131: 2007-2021.

38. Prokupek A, Hoffmann F, Eyun SI, Moriyama E, Zhou M, et al. (2008) An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. Evolution 62: 2936-2947.

39. Prokupek AM, Eyun SI, Ko L, Moriyama EN, Harshman LG (2010) Molecular evolutionary analysis of seminal receptacle sperm storage organ genes of *Drosophila melanogaster*. Journal of Evolutionary Biology 23: 1386-1398.

40. Prokupek AM, Kachman SD, Ladunga I, Harshman LG (2009) Transcriptional profiling of the sperm storage organs of *Drosophila melanogaster*. Insect Molecular Biology 18: 465-475.

41. Clark NL, Alani E, Aquadro CF (2012) Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Research 22: 714-720.

42. Drummond DA, Raval A, Wilke CO (2006) A single determinant dominates the rate of yeast protein evolution. Molecular Biology and Evolution 23: 327-337.

43. Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, et al. (2008) Evolution of protein-coding genes in *Drosophila*. Trends in Genetics 24: 114-123.

44. Liao B-Y, Scott NM, Zhang J (2006) Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. Molecular Biology and Evolution 23: 2072-2080.

45. McInerney JO (2006) The causes of protein evolutionary rate variation. Trends in Ecology & Evolution 21: 230-232.

46. Mintseris J, Weng ZP (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America 102: 10930-10935.

47. Pal C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. Genetics 158: 927-931.

48. Pal C, Papp B, Lercher MJ (2006) An integrated view of protein evolution. Nature Reviews Genetics 7: 337-348.

49. Rocha EPC, Danchin A (2004) An analysis of determinants of amino acids substitution rates in bacterial proteins. Molecular Biology and Evolution 21: 108-116.

50. Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, et al. (2009) Coevolution of interacting fertilization proteins. Plos Genetics 5: e1000570.

51. Hakes L, Lovell SC, Oliver SG, Robertson DL (2007) Specificity in protein interactions and its relationship with sequence diversity and coevolution. Proceedings of the National Academy of Sciences of the United States of America 104: 7999-8004.

52. Kann MG, Shoemaker BA, Panchenko AR, Przytycka TM (2009) Correlated evolution of interacting proteins: Looking behind the mirrortree. Journal of Molecular Biology 385: 91-98.

53. Lovell SC, Robertson DL (2010) An integrated view of molecular coevolution in protein-protein interactions. Molecular Biology and Evolution 27: 2567-2575.

54. Clark NL, Alani E, Aquadro CF (2013) Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. Genetics 193: 529-538.

55. Clark NL, Aquadro CF (2010) A novel method to detect proteins evolving at correlated rates: identifying new functional relationships between coevolving proteins. Molecular Biology and Evolution 27: 1152-1161.

56. Goh CS, Cohen FE (2002) Co-evolutionary analysis reveals insights into protein-protein interactions. Journal of Molecular Biology 324: 177-192.

57. Pazos F, Valencia A (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. Protein Engineering 14: 609-614.

58. Sato T, Yamanishi Y, Kanehisa M, Toh H (2005) The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. Bioinformatics 21: 3482-3489.

59. Juan D, Pazos F, Valencia A (2008) High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. Proceedings of the National Academy of Sciences of the United States of America 105: 934-939.

60. Tabach Y, Billi AC, Hayes GD, Newman MA, Zuk O, et al. (2013) Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. Nature 493: 694-698.

61. Consortium DG (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203-218.

62. Kim YJ, Bartalska K, Audsley N, Yamanaka N, Yapici N, et al. (2010) MIPs are ancestral ligands for the sex peptide receptor. Proceedings of the National Academy of Sciences of the United States of America 107: 6520-6525.

63. Poels J, Van Loy T, Vandersmissen HP, Van Hiel B, Van Soest S, et al. (2010) Myoinhibiting peptides are the ancestral ligands of the promiscuous *Drosophila* sex peptide receptor. Cellular and Molecular Life Sciences 67: 3511-3522.

64. Chintapalli VR, Wang J, Dow JAT (2007) Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nature Genetics 39: 715-720.

65. Dean MD, Clark NL, Findlay GD, Karn RC, Yi XH, et al. (2009) Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). Molecular Biology and Evolution 26: 1733-1743.

66. Pilch B, Mann M (2005) Large scale proteomic analysis of human seminal plasma. Molecular & Cellular Proteomics 4: S205-S205.

67. Sirot LK, Hardstone MC, Helinski MEH, Ribeiro JMC, Kimura M, et al. (2011) Towards a semen proteome of the dengue vector mosquito: Protein identification and potential functions. Plos Neglected Tropical Diseases 5.

68. Ross J, Jiang H, Kanost MR, Wang Y (2003) Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. Gene 304: 117-131.

69. Carpentier M, Guillemette C, Bailey JL, Boileau G, Jeannotte L, et al. (2004) Reduced fertility in male mice deficient in the zinc metallopeptidase NL1. Molecular and Cellular Biology 24: 4428-4437.

70. Pintado CO, Pinto FM, Pennefather JN, Hidalgo A, Baamonde A, et al. (2003) A role for tachykinins in female mouse and rat reproductive function. Biology of Reproduction 69: 940-946.

71. Pinto FM, Armesto CP, Magraner J, Trujillo M, Martin JD, et al. (1999) Tachykinin receptor and neutral endopeptidase gene expression in the rat uterus: Characterization and regulation in response to ovarian steroid treatment. Endocrinology 140: 2526-2532.

72. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8: 785-786.

73. Thomas JE, Rylett CM, Carhan A, Bland ND, Bingham RJ, et al. (2005) Drosophila melanogaster NEP2 is a new soluble member of the neprilysin family of endopeptidases with implications for reproduction and renal function. Biochem J 386: 357-366.

74. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research 39: D225-D229.

75. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nature Protocols 4: 363-371.

76. LaFlamme BA, Wolfner MF (2013) Identification and function of proteolysis regulators in seminal fluid. Molecular Reproduction and Development 80: 80-101.

77. Ravi Ram K, Sirot LK, Wolfner MF (2006) Predicted seminal astacin-like protease is required for processing of reproductive proteins in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America 103: 18674-18679.

78. Smith JR, Stanfield GM (2011) TRY-5 Is a sperm-activating protease in *Caenorhabditis elegans* aeminal fluid. Plos Genetics 7.

79. Stanfield GM, Villeneuve AM (2006) Regulation of sperm activation by SWM-1 is required for reproductive success of *C. elegans* males. Current Biology 16: 252-263.

80. Zhao Y, Sun W, Zhang P, Chi H, Zhang M-J, et al. (2012) Nematode sperm maturation triggered by protease involves sperm-secreted serine protease inhibitor (Serpin). Proceedings of the National Academy of Sciences of the United States of America 109: 1542-1547.

81. Chen PS, Stumm-Zollinger E, Aigaki T, Balmer J, Bienz M, et al. (1988) A male accessory gland peptide that regulates reproductive behavior of female *Drosophila melanogaster*. Cell 54: 291-298.

82. Hanin O, Azrielli A, Zakin V, Applebaum S, Rafaeli A (2011) Identification and differential expression of a sex-peptide receptor in *Helicoverpa armigera*. Insect Biochemistry and Molecular Biology 41: 537-544.

83. Dottorini T, Nicolaides L, Ranson H, Rogers DW, Crisanti A, et al. (2007) A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior. Proceedings of the National Academy of Sciences of the United States of America 104: 16215-16220.

84. Markow TA (1996) Evolution of *Drosophila* mating systems. Evolutionary Biology, Vol 29 29: 73-106.

85. Thailayil J, Magnusson K, Godfray HCJ, Crisanti A, Catteruccia F (2011) Spermless males elicit large-scale female responses to mating in the malaria mosquito *Anopheles gambiae*. Proceedings of the National Academy of Sciences of the United States of America 108: 13677-13681.

86. Clark NL, Aagaard JE, Swanson WJ (2006) Evolution of reproductive proteins from animals and plants. Reproduction 131: 11-22.

87. Findlay GD, Swanson WJ (2010) Proteomics enhances evolutionary and functional analysis of reproductive proteins. BioEssays 32: 26-36.

88. Turner LM, Hoekstra HE (2008) Causes and consequences of the evolution of reproductive proteins. International Journal of Developmental Biology 52: 769-780.

89. Kuijper B, Stewart AD, Rice WR (2006) The cost of mating rises nonlinearly with copulation frequency in a laboratory population of *Drosophila melanogaster*. Journal of Evolutionary Biology 19: 1795-1802.

90. Rice WR (1996) Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature 381: 232-234.

91. Edward DA, Fricke C, Gerrard DT, Chapman T (2011) Quantifying the life-history response to increased male exposure in female *Drosophila melanogaster*. Evolution 65: 564-573.

92. Wigby S, Chapman T (2005) Sex peptide causes mating costs in female *Drosophila melanogaster*. Current Biology 15: 316-321.

93. Cirera S, Aguade M (1997) Evolutionary history of the sex peptide (Acp70A) gene region in *Drosophila melanogaster*. Genetics 147: 189-197.

94. Wong A, Turchin MC, Wolfner MF, Aquadro CF (2008) Evidence for positive selection on Drosophila melanogaster seminal fluid protease homologs. Molecular Biology and Evolution 25: 497-506.

95. Arnone JT, Robbins-Pianka A, Arace JR, Kass-Gergi S, McAlear MA (2012) The adjacent positioning of co-regulated gene pairs is widely conserved across eukaryotes. BMC Genomics 13.

96. Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, et al. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Research 38: D196-D203.

97. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

98. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586-1591.

99. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, et al. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature 448: 151-U151.

100. Ni J-Q, Zhou R, Czech B, Liu L-P, Holderbaum L, et al. (2011) A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. Nature Methods 8: 405-U446.

101. Ni J-Q, Liu L-P, Binari R, Hardy R, Shim H-S, et al. (2009) A *Drosophila* resource of transgenic RNAi lines for neurogenetics. Genetics 182: 1089-1100.

102. Ni J-Q, Markstein M, Binari R, Pfeiffer B, Liu L-P, et al. (2008) Vector and parameters for targeted transgenic RNA interference in *Drosophila melanogaster*. Nature Methods 5: 49-51.

103. Xue L, Noll M (2002) Dual role of the Pax gene paired in accessory gland development of *Drosophila*. Development 129: 339-346.

104. Schnakenberg SL, Matias WR, Siegal ML (2011) Sperm-storage defects and live birth in *Drosophila* females lacking spermathecal secretory cells. Plos Biology 9.

CHAPTER 5

A NOVEL FUNCTION FOR THE HOX GENE *ABD-B* IN THE MALE ACCESSORY GLAND REGULATES THE LONG TERM FEMALE POST-MATING RESPONSE IN *DROSOPHILA*[5]

5.1 INTRODUCTION

The homeotic transcription factor *Abdominal-B* (*Abd-B*) specifies the identity of the four most-posterior abdominal segments of the fly (the 5th through 8th abdominal segments), as well as the genital and anal structures [1,2,3]. Each of these segments is specified by a particular pattern and level of Abd-B protein expression in the early embryo. Four segment-specific *cis*-regulatory domains (*iab-5* through *iab-8*) spanning >90kb of DNA have been shown to control the expression pattern of *Abd-B,* where each domain is predominantly responsible for controlling the *Abd-B* expression pattern in one particular segment [4,5,6] (for a review see [7]).

Extensive study has been devoted to exploring how the segment-specific expression pattern of *Abd-B* is achieved. Due to the striking cuticular transformations elicited by *Abd-B* mutations, genetic and transgenic analyses have been able to discover numerous enhancers, silencers and insulators that direct *Abd-B* expression in the ectoderm [8,9,10,11,12,13,14,15,16]. However, much less is known about the role of *Abd-B* in non-ectodermally derived tissues during later stages of development. Here, we use a 111kb BAC-based reporter construct to identify new locations of *Abd-B* expression in the adult fly. We find that *Abd-B* is strongly

[5] This chapter was published in PLoS Genetics as Gligorov, D.*, Sitnik, J.L.*, Maeda, R.K., Wolfner, M.W., and Karch, F., (2013) "A novel function for the hox gene *Abd-B* in the male accessory gland regulates the long-term female post-mating response in *Drosophila.*"(*authors contributed equally). Characterization of the *iab-6*^{cocu} mutant and *Abd-B* expression conducted by Dragan Gligorov is acknowledged in the figure legends. Supplementary material not included in this chapter can be found elextronically: at doi: 10.1371/journal.pgen.1003395. Reprinted with permission.

expressed in the accessory gland (AG), a secretory tissue of the adult male reproductive tract that has important reproductive functions.

The AG synthesizes seminal proteins that are essential for male fertility. These >180 accessory gland proteins ("Acps") are transferred to females during mating and cause post-mating changes in the females known collectively as the post-mating response (PMR). The PMR includes increased rates of egg-laying and ovulation, sperm storage, decreased receptivity to courting males, as well as changes in longevity, feeding, and sleep patterns (reviewed in [17,18]). The PMR is divided into two phases. The short term response (STR) refers to changes in the above behaviors during the first ~24 hours post-mating. The STR requires Acps, but not the receipt of sperm. Persistence of the PMR after 24 hr (and for up to ~10 days) is known as the long-term response (LTR). The LTR requires Acps and stored sperm [19,20,21,22]. Many of the roles of Acps were initially discovered by experiments in which whole AG extracts or purified Acps were injected into unmated females [23,24,25], or by whole-tissue ablation in males [26].

Each lobe of the AG is composed of a monolayer of approximately 1000 secretory cells comprised of two morphologically distinct cell types. Roughly 96% of these cells are flat, polygonally shaped "main cells". The remaining 4% of the cells are large, spherical, vacuole filled "secondary cells "; these are dispersed among the main cells at the distal tip of the gland. Enhancer trapping and other studies have shown that, in addition to their morphological differences, these two secretory cell types are biochemically distinct [27,28,29]. Ablation of the main cells only [19] showed that products of these cells are essential for the PMR. These products include ovulin (Acp26Aa), an Acp that acts in the STR to stimulate ovulation [30,31], and the sex peptide (SP, Acp70A), which is the ultimate regulator of most other PMR effects [22,32,33,34,35]. SP binds to sperm within the mated female, and its active portion is gradually

released from the sperm [22]. This binding and release allows SP to affect the female for as long as she contains stored sperm. A network of five other Acps is necessary for SP to bind to sperm and enter storage. The predicted protease CG10586 (Seminase) [36] appears to be necessary for both STR and LTR related events, while the predicted protease CG9997, the predicted cysteine-rich secretory protein (CRISP) CG17575, and the predicted lectins CG1656/1652 appear to be LTR specific [37,38,39,40]. The cellular source of each of these proteins is currently unknown.

In spite of the detailed characterization of the main cells and several specific Acps, the role of the secondary cells has remained mysterious. No PMR-associated Acps were known to be expressed exclusively in the secondary cells, and no tools have been available to specifically target those cells. Here, we identified the secondary cells of the male AG as a novel location of *Abd-B* expression in the adult fly. By screening an extensive collection of *cis*-regulatory deletions [6,41,42], we discovered a 2.8kb enhancer from the *iab-6 cis*-regulatory domain, whose removal completely abolishes *Abd-B* expression in the secondary cells. Loss of *Abd-B* expression in the secondary cells causes those cells to develop aberrantly. Moreover, these mutant males provide their mates with substances that initiate the PMR, but are insufficient to maintain it. Our results indicate that *Abd-B* expression in the secondary cells is essential for their proper development and for the production of proteins important for long-term changes in female post-mating responses.

5.2 **RESULTS**

Creation of Gal4 reporter BAC for Abdominal B.

In order to discover new tissues in which the *Abd-B* gene functions, we undertook the creation of a transgenic reporter that accurately reproduces the *Abd-B* expression pattern

throughout development. Previous studies indicated that the *Abd-B* gene is expressed as two isoforms, the Abd-B *m* and *r* forms, and that the expression of these two isoforms requires separate elements located within a large *cis*-regulatory region spanning >90kb of DNA [43]. As the *Abd-Br* isoform is thought to be primarily involved in the formation of the external genitalia [44], we decided to concentrate our study on the *Abd-Bm* isoform, which is involved in determining segment identity. BACR24L18 is a BAC of ~172kb that contains the *Abd-B,* and much of the *abd-A* region of the Bithorax complex (BX-C). By recombineering, we reduced BAC24L18 to contain mostly the *iab-5* to *iab-8* domains required for *Abd-Bm* expression (removing many of the *Abd-Br* alternative promoters and its regulatory elements) and the *Abd-Bm* coding sequence (Figure 5.1B). A ΦC31 AttB integration sequence and a *white* integration marker were also added during the reduction step (Figure 5.1B&C).

We first tested if expression derived from the sequences on this BAC were sufficient to rescue *Abd-B* mutant phenotypes. We integrated the *Abd-B* BAC into the 51C landing platform [45] and tested for complementation of two large deletions affecting *Abd-B* activity. We found that the presence of a copy of the BAC on the second chromosome rescues the mutant phenotypes of *iab*-6,7^{IH} and *iab*-5,6^{J82} [6] (Supplementary, Figure S1[46]). Because the sequences preserved on the BAC seemed to drive appropriate *Abd-Bm* expression, we proceeded to modify the BAC by recombineering to replace the first codon of the first exon of *Abd-B* with the sequence encoding the Gal4 transcription factor. As this sequence also adds a stop codon, the expression of *Abd-B* from the BAC should be eliminated, but any sequences that might be used in *Abd-B* gene regulation will be preserved to drive reporter gene expression. The final BAC used in the experiments was 111kb (Fig.5.1). It was integrated into the 51C landing platform.

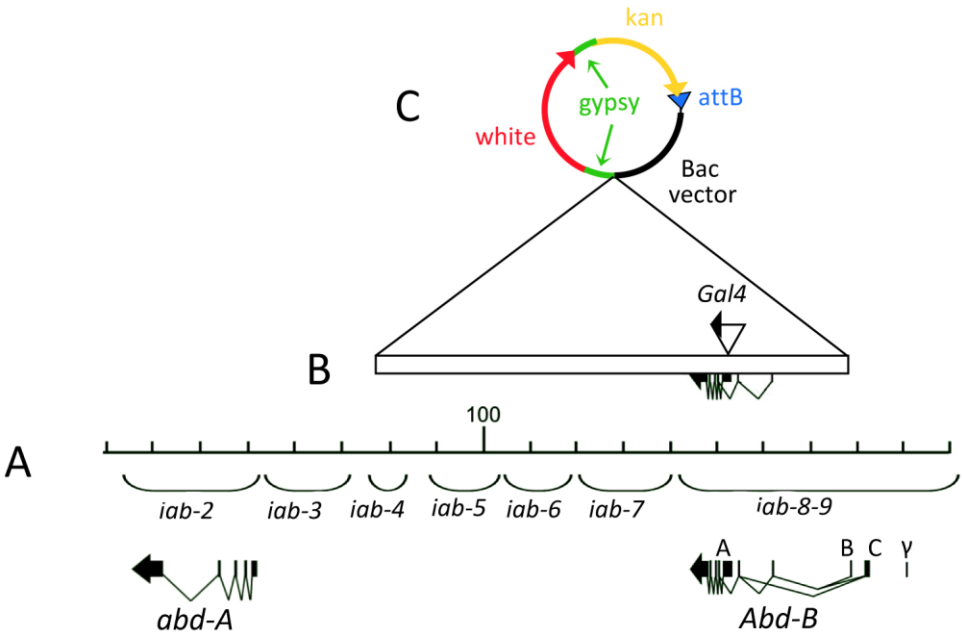


Figure 5.1: Extent of DNA contained in the *Abd-B* BAC

A) Molecular map of the abdominal region of the Bithorax complex numbered in kb according to [47] (Genbank U31961). The *abd-A* and *Abd-B* transcription units are drawn below the DNA line along with the extent of the segment-specific *iab cis*-regulatory domains *iab-2* through *iab-9*. **B)** The rectangle depicts the extent of the BAC used in this study. Note that it lacks the B,C and γ promoters specific for the *Abd-Br* form. The Gal-4 coding sequence was inserted within the 5'UTR of the *Abd-Bm* form. **C)** The structure of the vector sequences used to propagate the BAC and to select the integration within the *Drosophila* genome. Note the presence of two gypsy insulator sequences flanking the mini-*white* sequences to prevent possible position effect on *white* expression (see material and methods for further details). Work from D. Gligorov and R. Maeda.

To study the *Abd-B* expression pattern, a line was established containing the *Abd-B-Gal4* BAC and a UAS-GFP reporter. Initial examination of the embryonic expression pattern in these lines confirms that the *Abd-B-Gal4* BAC appears to recapitulate most of the wild-type expression pattern of *Abd-Bm* in early embryos (Figure 5.2A; Figure.S2 and Figure.S3 [46]). Later, we do observe some evidence of ectopic expression from the BAC, particularly in the ventral nerve cord (Figure S2; Figure S3[46]). Even with the slight level of ectopic expression, the *Abd-B-Gal4 BAC* seems to be a useful tool, as it recapitulates known patterns of *Abd-B* expression even in adult and larval tissues (Figure S4[46]).

Previously unknown location of Abdominal B expression in adult flies.

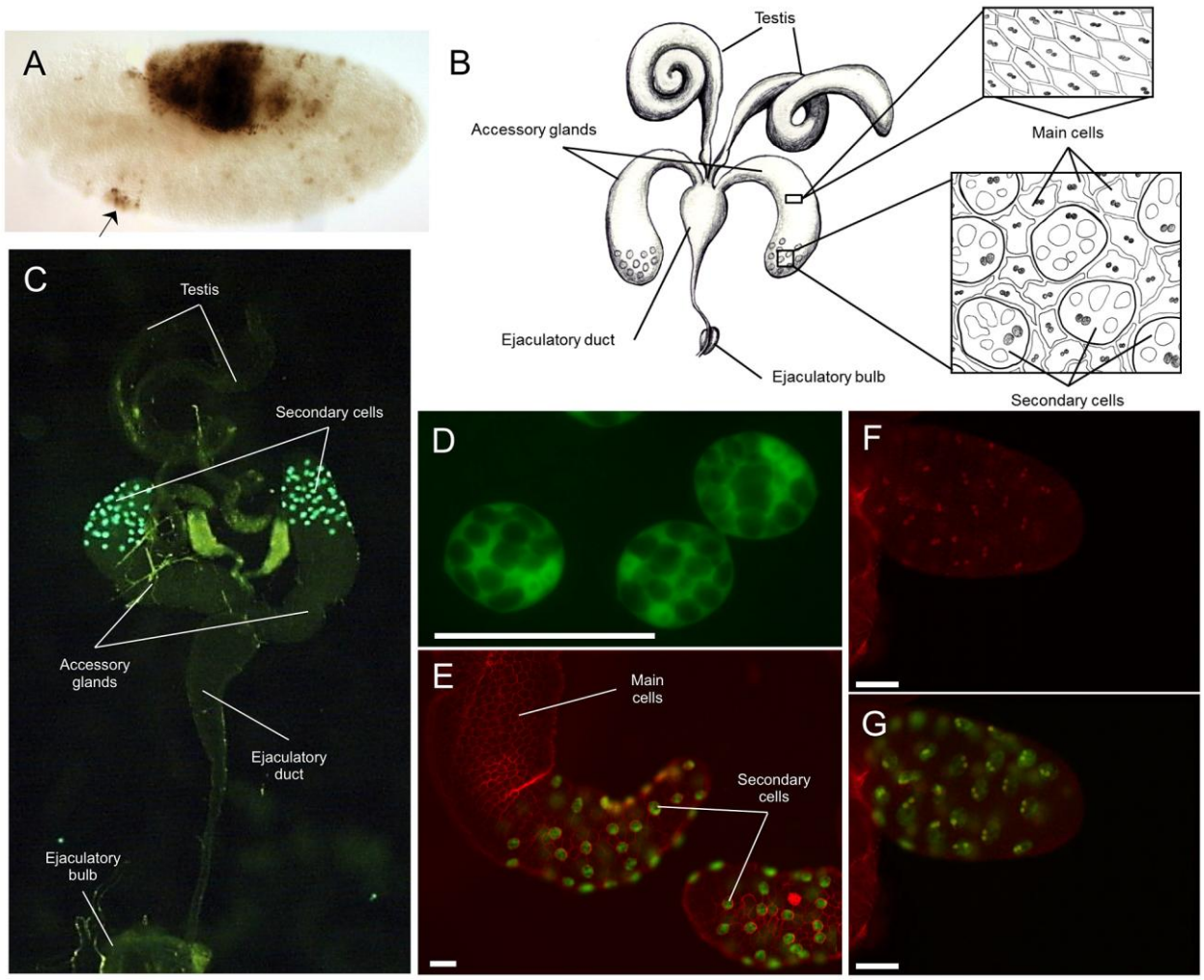
Using this new reporter, we identified the adult male accessory gland (see Figre 5.2B) as a location of *Abd-B* expression (Figure 5.2C). More specifically, based on the expression of our *Abd-B-Gal4* BAC, *Abd-B* appears to be specifically expressed in the secondary cells (Figure 5.2C,D,&E). To confirm this finding, we stained accessory glands in the presence of the *Abd-B-Gal4 UAS-GFP* reporter with an antibody directed against *Abd-B* (Figure 5.2F&G). Like the reporter, accessory gland immunostaining against the *Abd-B* protein shows specific staining in the secondary cells (Figure 5.2F&G). Interestingly, we also see *Abd-B* staining in the ejaculatory duct (Fig. S4C) that is not observed with our reporter (Figure 5.2C). This is perhaps not surprising, as the ejaculatory duct is a structure derived from the male genital disc, a tissue that primarily expresses the *Abd-Br* isoform [44] (which is also recognized by our antibody).

Secondary cell enhancer.

In order to examine the role of *Abd-B* in the development of the accessory glands, we

Figure 5.2: Expression patterns driven by the *Abd-B-Gal4* BAC

A) Embryo expressing the *Abd-B-Gal4* BAC crossed to a *UAS-LacZ* reporter stained with an antibody directed against ß-galactosidase. Out of a slight ectopic expression anteriorly (indicated by the arrow), the expression pattern is stickingly simiar to the WT *Abd-B* expression pattern as documented in Figure S2. **B)** Cartoon depicting the male reproductive apparatus with testis, the paired accessory glands, the ejaculatory duct and ejaculatory bulb. Each accessory gland contains two secretory cell types, the main cells which make up the majority of the gland (top insert) and the secondary cells which are located at the distal tip of the gland interspersed among the main cells (bottom insert) Drawing by J. L. Sitnik; **C)** Picture of the male reproductive system from flies carrying the *Abd-B-Gal4* BAC crossed to a *UAS-GFP* reporter with the secondary cells of the accessory glands showing GFP expression. The different organs composing the system are marked. **D)** Magnification of three secondary cells from flies carrying the *Abd-B-Gal4* BAC crossed to a cytoplasmic *UAS-GFP reporter*. The multiple, large vacuoles, characteristic of secondary cells, can be visualized through their exclusion of the GFP protein. The two nuclei of the cells can also be seen as slightly more intense GFP signals..; **E)** The tip of the accessory gland with GFP expressed specifically in the secondary cells driven by the *Abd-B-Gal4* BAC (green), co-stained with the membrane staining dye, FM4-64, in red. The two cell types can be clearly distinguished with examples indicated with white lines.; **F)** Abd-B antibody staining of the tip of an accessory gland on *Abd-B-Gal4* BAC, *UAS-GFP* flies (red). Only secondary cell nuclei are stained. **G)** GFP expression (green) in the same gland overlaid onto the *Abd-B* antibody staining (red) shown in Figure 2F. Each cell with Abd-B protein expression also express GFP. The white scale bar on figures D, F, E, and G represents 50μm. Data from D. Gligorov.

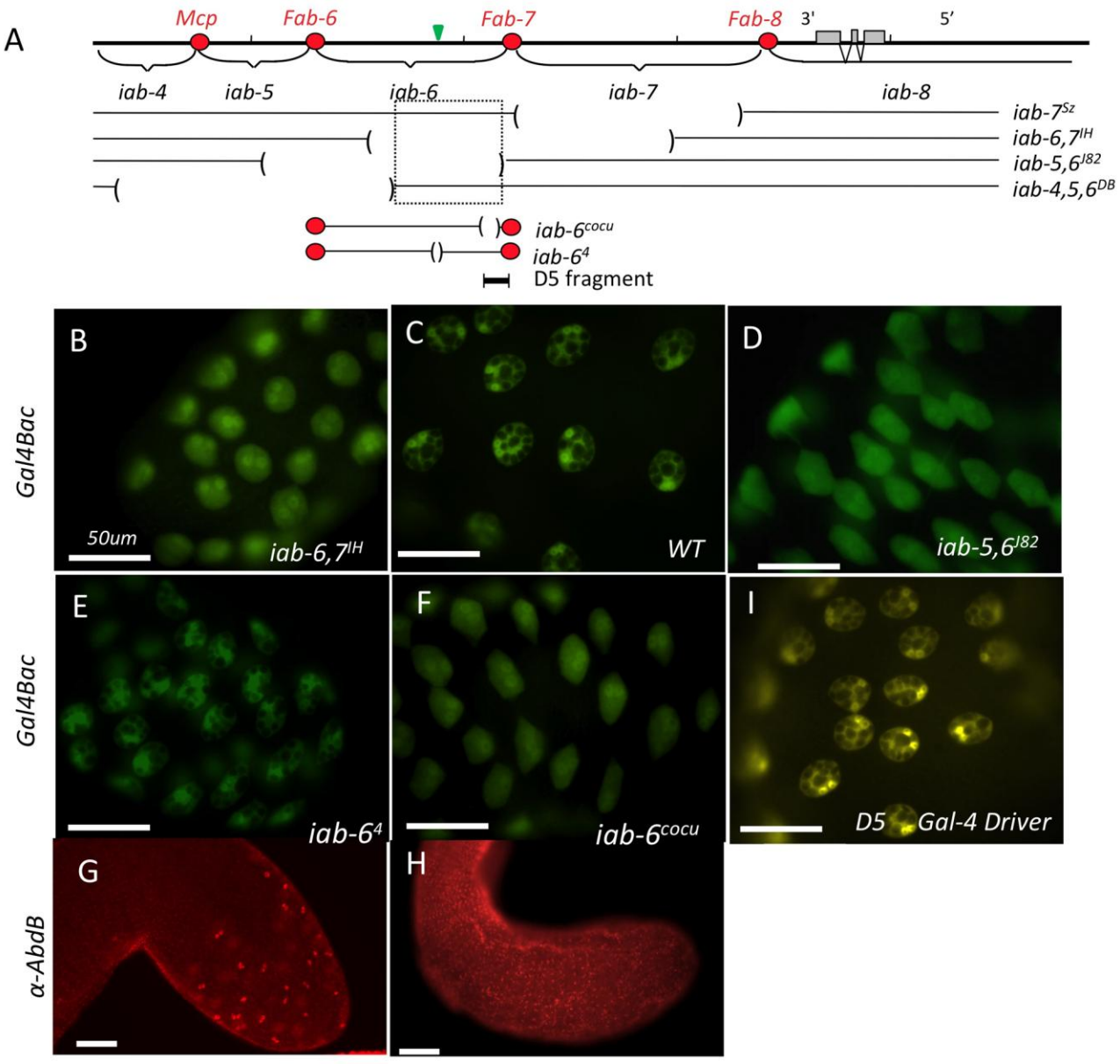

A
B
Testis
Accessory glands
Main cells
Ejaculatory duct
Ejaculatory bulb
Secondary cells
C
Testis
Secondary cells
Accessory glands
Ejaculatory duct
Ejaculatory bulb
D
E
Main cells
Secondary cells
F
G

sought a method to remove *Abd-B* expression exclusively in the secondary cells. Rather than use the traditional FLP-FRT system for making clones, we reasoned that in our collection of *Abd-B* *cis*-regulatory mutations [6], we may already have a deletion that specifically removes secondary cell enhancers. Given our hypothesis that *Abd-B* might act as a cell fate determinant in the secondary cells, we screened a set of large, overlapping deficiencies covering the *Abd-B cis*-regulatory region for defects in secondary cell formation (Figure 5.3A). To make this analysis easier, homozygous mutant flies were screened in lines that also contain a copy of our BAC reporter to mark the cells that would normally become secondary cells. Two of the lines examined, *iab*-6,7^{IH} & *iab*-5,6^{J82} (Figure 5.3C&D), showed a distinct morphological abnormality in the secondary cells. This abnormality can be easily seen using the cytoplasmic GFP marker. In wild-type cells, the GFP marker outlines the presence of large vacuolar structures in the secondary cells (Figure 5.2D, 5.3B). In both the *iab*-6,7^{IH} & *iab*-5,6^{J82} mutants, these structures appear to be absent, and consequently the GFP marker is almost uniformly distributed across the cytoplasm.

Although these secondary cells are not normal, we do not detect any expression of main cell-specific markers in these cells, suggesting that they are not transformed towards a main cell fate (they still express the Acp95EF lacZ reporter gene [28] and fail to express the SP lacZ reporter gene (data not shown) [29]). To test if *Abd-B* is capable of transforming main cells into secondary cells, we expressed *Abd-B* across the whole accessory gland using a *paired-Gal4* driver [48]. The most common result of this ectopic expression is cell death in the main cells (data not shown). These results suggest that *Abd-B* expression in the secondary cells is required for morphological differentiation but may not be necessary for the initial differentiation between the two cell types.

Figure 5.3: Mutants affecting *Abd-B* expression in the accessory gland

A) The Molecular map of the *Abd-B* gene region is shown with its extensive 3' *cis*-regulatory domains *iab-5* through *iab-8* (the *iab-4* domain regulates *abd-A*). The extents of the various deficiencies that were used to map the enhancer responsible for *Abd-B* expression in the secondary cells are shown below the molecular map. The location of DNA sequence used to make the 2.8kb-long D5rsG4rs driver (thereby refereed as D5 Gal4 driver).is shown under the map. The red circles on the map represent the boundaries separating the parasegment-specific *cis*-regulatory domains of *Abd-B*. The green triangle above the *iab-6* domain marks the *iab-6* initiator. **B)** UAS-GFP expression driven by *Abd-B*-Gal4 in a WT for the BX-C. **C)** same as B**,** but in an *iab-6, 7*^{IH} homozygous male or in an *iab-5,6*^{J82} homozygous male(**D**). Note that in the *iab-6, 7*^{IH} and *iab-5,6*^{J82} background, the numerous vacuoles, characteristic of the secondary cells (visible by black holes in the GFP background), are lost. However, the vacuoles are not affected in *iab-4,5,6*^{DB}. Thus, the critical region required for proper secondary cell specification based on these 3 deficiencies is indicated by the dotted-line box in panel A. **E)** UAS-GFP expression driven by *Abd-B*-Gal4 in secondary cells of *iab-6*^{4} (initiator deletion) and of *iab-6*^{cocu} males (**F**). Note the normal aspect of GFP staining in *iab-6*^{4} (E) relative to the WT shown in B). In *iab-6*^{cocu} however (F), the vacuoles are lost, giving rise to staining comparable to panels C and D. Panels **G)** and **H)** show *iab-6*^{4} (G) and *iab-6*^{cocu} (H) accessory glands stained with an *Abd-B* antibody. While *Abd-B* expression appears normal in *iab-6*^{4} (G), the signal is absent in *iab-6*^{cocu} (H). **I)** shows the tip of an accessory gland from a fly carrying the D5-Gal4 driver driving GFP expression in the secondary cells (the staining is shown in yellow to distinguish it from panels B-F depicting GFP driven by the Gal4 Bac. The white horizontal scale bars in each of the panels represents 50um. Data from D. Gligorov.


A
Mcp
Fab-6
Fab-7
Fab-8
3'
5'
iab-4
iab-5
iab-6
iab-7
iab-8
iab-7Sz
iab-6,7IH
iab-5,6J82
iab-4,5,6DB
iab-6cocu
iab-64
D5 fragment
Gal4Bac
B
50um
iab-6,7IH
C
WT
D
iab-5,6J82
Gal4Bac
E
iab-64
F
iab-6cocu
I
D5 Gal-4 Driver
α-AbdB
G
H

Based on the sequences uncovered by both the *iab*-6,7^{IH} & *iab*-5,6^{J82} mutations, we concluded that the *iab-6* domain, responsible for *Abd-B* expression in segment 6, is also responsible for *Abd-B* expression in the secondary cells. Thus, we screened our collection of smaller *iab-6* deficiencies [41] for the secondary cell phenotype. From this analysis, we were able to narrow down the location containing the secondary cell enhancer to a 2.8kb region in *iab*-6 (Figure 5.3A,F,&H). Flies lacking this 2.8 kb region (*iab-6*$^{\Delta 5}$) specifically lack Abd-B protein expression in the secondary cells (Figure 5.3H), and show distinct secondary cell morphological defects (Figure 5.3F). Like the larger deficiencies above, *iab-6*$^{\Delta 5}$ homozygous males lack the large vacuoles characteristic of secondary cells.

As further confirmation of the importance of *Abd-B* and the 2.8 kb *iab-6* enhancer in secondary cell development, we performed a number of control experiments. First, we crossed in a BAC transgene containing the wild-type *Abd-B* region and tested for rescue of the cellular phenotype. As expected, the secondary cells of males, homozygous for the *iab-6*$^{\Delta 5}$ mutation but carrying one copy of the *Abd-B* BAC are substantially rescued (containing a number of large vacuoles) (Supplementary, Figure S5C[46]). Although this rescue is quite evident, it is not complete, a fact that probably reflects a weaker level of expression from the BAC relative to the native *Abd-B* locus. Indeed, *Abd-B* staining experiments using this BAC indicate that this is the case (data not shown). Next, we created a transgene carrying the 2.8 kb region of *iab-6* (called D5-Gal4) and showed that it drives expression of Gal4 in the male reproductive tract specifically in the secondary cells (Figure 5.3E). Using this D5-Gal4 driver, we were then able to drive expression of an Abd-B RNAi construct in the secondary cells. Knocking down Abd-B in the secondary cells was able to partially phenocopy the *iab-6*$^{\Delta 5}$ mutation (Supplementary Figure

S6C&F [46]). The strength of this phenotype could be enhanced by the inclusion of a Dicer 2 overexpression transgene in the background.

iab-6$^{\Delta 5}$ was originally isolated in Iampietro et al. [41], where they did not observe any visible external phenotype. With the discovery of the secondary cell phenotype and the strong reproductive phenotype described below, we have renamed this allele *iab-6*^{cocu} ("cocu" means "cuckold" in French, reflecting that the mates of these males fail to reject other suitors).

Abd-B expression in secondary cells is independent of the initiator.

Interestingly, although the secondary cell enhancer was found in the *iab-6* domain, it does not seem to be regulated like other BX-C enhancers. Previous work has demonstrated that most enhancers in the BX-C function coordinately through their integration into segment-specifically activated chromatin domains [6,49,50]. A special domain control element, called an initiator, is thought to dictate the activity state of a domain along the A-P axis [41,51,52]. Thus, deletion of the *iab-6* initiator is predicted to inactivate *Abd-B* expression in the secondary cells, because the secondary cell enhancer should be coordinately regulated with the other enhancers in the *iab-6* domain. In contrast to this prediction, we observed that deletions of the *iab-6* initiator, which seem to show complete transformations of A6 to A5, display wild-type accessory glands (Figure 5.3E&G). From these experiments, we conclude that *Abd-B* expression in the secondary cells is set up by a different mechanism than that of tissues arising early in development.

Impact of iab-6^{cocu} on the production of main cell Acps.

The *iab-6*^{cocu} mutation offers the opportunity to investigate the role of the secondary cells in the PMR. First, we tested whether the main cells of these males are functional, since loss of main

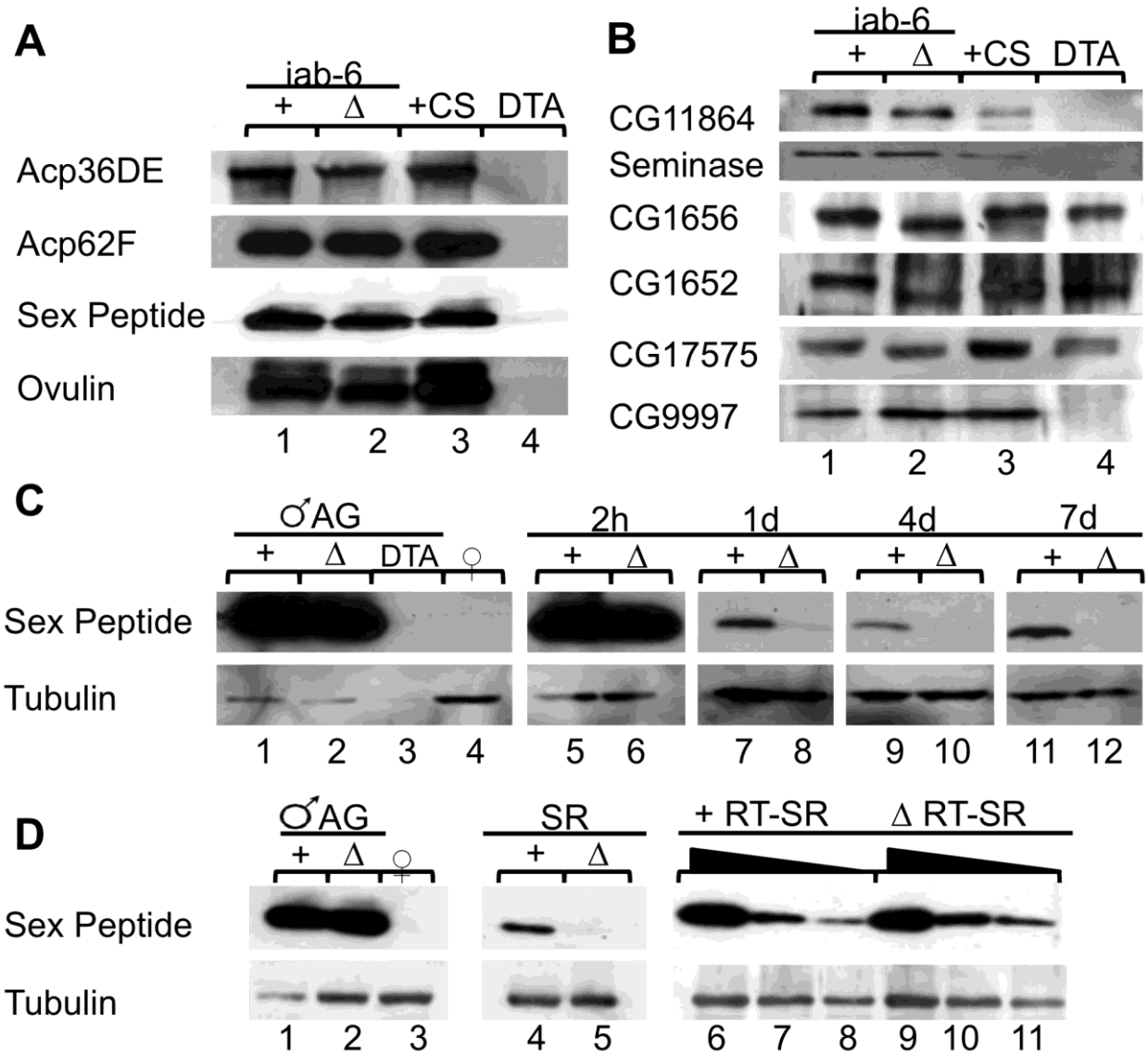
cell derived Acps may mask any secondary cell related phenotypes present in our mutant. We performed Western blots to examine the presence of known main cell Acps in the accessory glands of *iab-6*^{cocu} males relative to two types of control males [males heterozygous for the *iab-6*^{cocu} mutation (henceforth referred to as control males) and wild type males (Canton S)]. As a negative control, we included the accessory glands of DTA-E males, which lack protein production in the main cells [19] but have apparently normal secondary cells. We used antibodies to four Acps expressed in the main cells: SP [29], Acp62F, Acp36DE [53], and ovulin; the latter Acp is also present in the secondary cells but is known to be absent in DTA-E males [54]. We detected all four Acps in the extracts from *iab-6*^{cocu} males (Figure.5.4A). This result suggests that the main cells in *iab-6*^{cocu} males are functional.

Egg laying in mates of iab-6^{cocu} males.

To test if the *iab-6*^{cocu} mutation impacts the ability of males to induce egg-laying in their mates, we crossed *iab-6*^{cocu} males, control males, and DTA-E males to virgin females. During the first 24 hours after the start of mating (ASM), the number of eggs laid by females that had mated to *iab-6*^{cocu} males is comparable to that of mates of control males and is significantly higher than the number laid by females mated to DTA-E males (Figure 5.5A). This indicates that the *iab-6*^{cocu} mutation does not impact the STR and supports the Western blot results that suggest that the main cells are normal. However, egg laying in mates of *iab-6*^{cocu} males decreased dramatically at 48 hours, and the total number of eggs produced over the entire 10 day period was significantly lower than the number laid by mates of control males (Fig. 4.5A) (note that Canton-S males behave similarly to our control males (Figure. S7 [46]). This drop in egg laying is consistent with that observed when females do not receive or fail to store/release SP.

Figure 5.4: Seminal fluid proteins in *iab-6^{cocu}* and control males, and their mates

A) Western blots of accessory gland extracts from two control males (lane 1), two *iab-6^{cocu}* males (lane 2), two wild type males (lane 3), and two DTA-E males (lane 4). All Acps known to be produced by the primary cell (Acp36DE, Acp62F, sex peptide, and ovulin) are present in the accessory glands of *iab-6^{cocu}* males, but not DTA-E males. **B)** Other Acps necessary for various aspects of the PMR (CG11864, Seminase, CG1656, CG1652, CG17575, and CG9997) are present in the accessory glands of *iab-6^{cocu}* males. CG1656, CG1652, and CG17575 are always detectable in DTA-E males, however their abundance is highly variable compared to controls (The western blots depicted were selectived to most clearly demonstrate the presence of these proteins in DTA-E males). **C)** Mates of *iab-6^{cocu}* males have less SP, as detected by antibodies to SP, in the reproductive tract at 1d ASM and all subsequent time points. Tubulin was used as a loading control for the female reproductive tracts. Accessory gland extracts from a single control male (lane 1) and *iab-6^{cocu}* male (lane 2) were used as a positive control and accessory gland extracts from 2 DTA-E males (lane 3) and reproductive tract extracts from 8 virgin females (lane 4) were used as a negative control. Reproductive tract extracts from females mated to either control (+) or *iab-6^{cocu}* (Δ) males at 2h (lane 5-6, 2 RTs per), 1d (lane 7-8, 20 RTs per), 4d (lane 9-10, 18 RTs per), and 7d ASM (lane 11-12, 21 RTs per). **D)** Mates of *iab-6^{cocu}* males have dramatically less SP in the seminal receptacle (SR) at 2h ASM. Tubulin was used as a loading control for the female reproductive tracts. Accessory gland extracts from a single control male (lane 1) and *iab-6^{cocu}* male (lane 2) were used as positive controls and reproductive tract extracts from 8 virgin females (lane 3) were used as a negative control. Extracts from SRs dissected from females mated to either control (+) or *iab-6^{cocu}* (Δ) males at 2h ASM (20 SRs each, lanes 4-5). The amount of SP present in the reproductive tract (minus the SR) of mates of control and *iab-6^{cocu}* males was determined in a dilution series (1:1 (lanes 6 and 9), 1:2 (lanes 7 and 10), and 1:4 (lanes 8 and 11) and are equivalent to 5 RTs, 2.5 RTs, and 1.25 RTs). There is no appreciable difference in the amount of transferred SP.


A
jab-6
+ Δ +CS DTA
Acp36DE
Acp62F
Sex Peptide
Ovulin
1 2 3 4
B
jab-6
+ Δ +CS DTA
CG11864
Seminase
CG1656
CG1652
CG17575
CG9997
1 2 3 4
C
♂AG
+ Δ DTA ♀
2h
1d
4d
7d
+ Δ
Sex Peptide
Tubulin
1 2 3 4 5 6 7 8 9 10 11 12
D
♂AG
+ Δ ♀
SR
+ Δ
+ RT-SR
Δ RT-SR
Sex Peptide
Tubulin
1 2 3 4 5 6 7 8 9 10 11

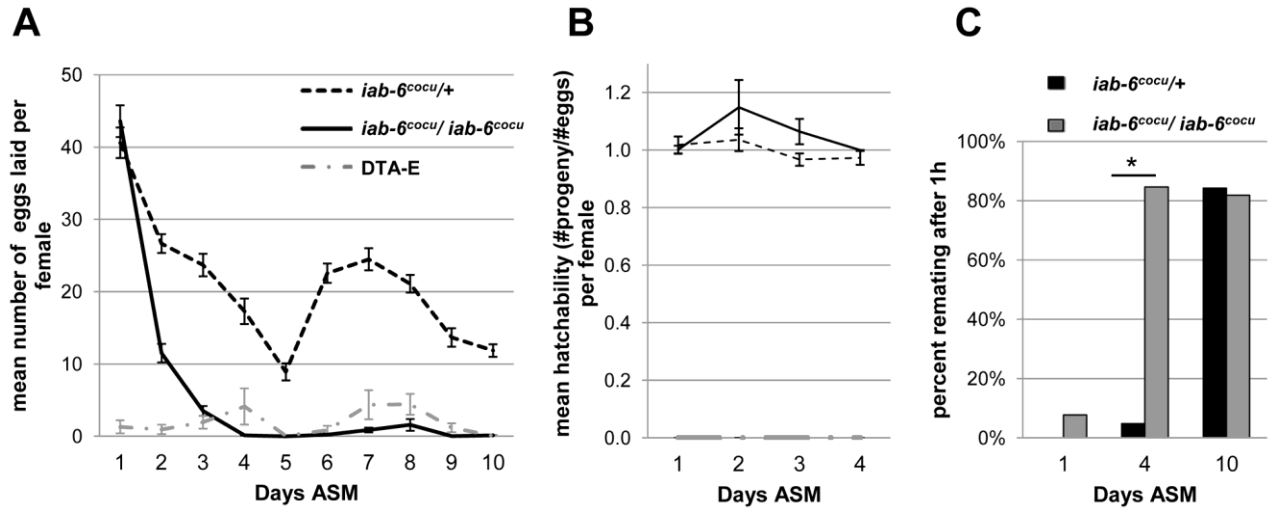


Figure 5.5: Egg-laying and receptivity in mates of *iab-6*^{cocu} or control males

A) The mean number of eggs laid per female mated to either control males (dashed line), *iab-6*^{cocu} males (solid line), or DTA-E males (grey dot dashed line) over a 10 day period. Mates of *iab-6*^{cocu} males lay normal numbers of eggs during the first day after mating (WRST p=0.300) but lay significantly fewer eggs over 10 days when compared to mates of control males (rmANOVA p= <0.0001*, Control N=51, *iab-6*^{cocu} N=45, DTA-E N=17). The drop in egg laying for controls on day 5 is atypical and was likely a response to food quality. **B)** The mean hatchability (#progeny/#eggs) per female for mates of control, *iab-6*^{cocu}, and DTA-E males for days 1-4 of the egg laying results reported in (A). Days 5-10 were omitted because *iab-6*^{cocu} mated females do not lay enough eggs on these days for analysis. Mates of *iab-6*^{cocu} males have comparable hatching totals for the eggs that they do lay when compared to mates of control males (WRST p=0.37). Because DTA-E males do not produce sperm, their mates are expected to show zero hatchability. (Values greater than 1 represent instances where the number of progeny produced exceeded the number of eggs counted; this under-counting can result when females lay eggs under bubbles in the medium or directly on top of previously laid eggs. Hatchability values were not normalized to 1 so as to accurately report counter error.) **C)** The percentage of mated females willing to mate within 1 hour of exposure to a wild type male at 1, 4, and 10 days after an initial mating. Both groups of females initially mated to *iab-6*^{cocu} or control males are unreceptive (WRST p=0.21, control n=20, *iab-6*^{cocu} n=26). At 4d ASM females initially mated to *iab-6*^{cocu} males are significantly more receptive to courting males compared to mates of control males (WRST p=<0.0001*, control n=21, *iab-6*^{cocu} n=26). By 10d ASM there is no difference between mates of control or *iab-6*^{cocu} males (WRST p=0.84, control n=19, *iab-6*^{cocu} n=22).

This suggests that products from the secondary cells may be necessary for maintenance of the LTR. The proportion of progeny that eclosed from eggs laid by females mated to either *iab-6*^{cocu} or control males was comparable, suggesting that there is no effect of secondary cell products on hatchability (Figure.5.5B). Together these results suggest that the secondary cells perform a function that is essential for the maintenance of the post-mating egg-laying increase, but does not impact hatchability, and that this function is perturbed in *iab-6*^{cocu} males.

Receptivity in mates of iab-6^{cocu} males.

Under normal conditions, mated females are less receptive to subsequent mating for more than four days after the initial mating occurred [55]. This reduction in receptivity requires the receipt of Acps [19,23,26]. To test whether the *iab-6*^{cocu} mutation alters female receptivity to remating, we mated virgin females to either *iab-6*^{cocu} or control males and then allowed these females access to a single WT male at 1d, 4d, or 10d ASM. At 24 hours after the initial mating, neither group of females remated, further suggesting that the STR is intact in mates of *iab-6*^{cocu} males. However, when mated females were introduced to a WT male at 4 days ASM, females which had initially mated to *iab-6*^{cocu} males were significantly more receptive than mates of control males. At 10 days ASM, both groups were fully receptive (Figure 5.5C). Our results show that sexual receptivity is initially repressed in mates of *iab-6*^{cocu} males, but that this effect is not maintained. This finding demonstrates a defect similar to those observed in known LTR-related proteins and further corroborates the LTR phenotype observed in our egg-laying experiments.

To verify that both the egg-laying and receptivity phenotypes are caused specifically by the loss of *Abd-B* expression in the secondary cells, we also performed these experiments using

BAC rescued *iab-6*^{cocu} flies (Supplementary Figure S5 [46]) and D5-Gal4::*Abd-B* RNAi flies (Supplementary Fig. S6 [46]). In both cases, these experiments confirm a role for *Abd-B* in producing the PMR phenotypes. Again, as with the cellular phenotypes mentioned above, both the rescue and phenocopying was incomplete, though clearly significant. Given the caveats involved in these experiments regarding the level and timing of protein expression, this was perhaps not unexpected and demonstrate a strong relationship between the celluar and the behavioral phenotypes. Nonetheless, these data clearly point to a major role for *Abd-B* expression in the secondary cells in modulating the PMR.

The production of LTR-associated Acps in iab-6^{cocu} males.

Our results suggest that the secondary cells are necessary for the processes required for long-term PMR maintenance. Therefore, the *iab-6*^{cocu} mutation likely impacts proteins required for the LTR. While the Acps that are produced and transferred to females have been extensively described, [53,56,57,58,59] little is known about their cellular origins. Thus, we investigated the possibility that some of the known PMR-related Acps, and more specifically those involved in the LTR, could be produced in the secondary cells or in both cell types, and thus, may be absent in *iab-6*^{cocu} males.

We performed Western blots to examine the presence of known LTR Acps in the accessory glands of *iab-6*^{cocu} males relative to control males. As a negative control for main cell expressed Acps, we included DTA-E males. Any secondary cell-expressed Acp should still be produced in these males, but main cell expressed Acps should not. We used antibodies to six Acps that regulate the PMR; one STR associated Acp (CG11864) and five LTR associated Acps (CG10586 (Seminase), CG9997, CG17575, CG1656, and CG1652 [36,38,39,40]). All of these

Acps were present in *iab-6*^{cocu} males (Figure 5.4B). Surprisingly, three of the Acps associated with the SP pathway (CG17575, CG1656, and CG1652) were also present in AG extracts from DTA-E males suggesting that they may be secondary cell expressed. Supporting this hypothesis, RNAi of these proteins in the secondary cells knocks down their expression, while leaving a main cell-expressed control protein, Acp62F, unchanged (Supplementary Figure S8 [46], see Chapter 6). The remainder of the Acps, CG9997, CG11864, and Seminase, are likely all expressed primarily or exclusively in the main cells. Further, these results support our previous conclusion that the secondary cells in *iab-6*^{cocu} males maintain a distinct expression profile from main cells and still produce some secondary cell proteins.

A role for secondary cells in SP storage.

The LTR defects seen in mates of *iab-6*^{cocu} males are consistent with those associated with failure to store or release SP [53,57,58,59]. However, *iab-6*^{cocu} males produce SP and all known LTR Acps. Still, it is possible that the *iab-6*^{cocu} mutation interferes with the ability of SP to enter storage and thus maintain the LTR. We tested for this by performing Western blots using SP antibodies. Both control and *iab-6*^{cocu} males transfer SP to their mates, and there are comparable amounts of SP in the female reproductive tract by 2h ASM. However, by 24 hours ASM and continuing to 6 days thereafter, mates of *iab-6*^{cocu} males have significantly less stored SP (Figure 5.4C; Figure.S9 [46]). To distinguish between premature loss of SP from the seminal receptacle (SR) versus failure of SP to be stored in the SR initially, we performed Western blots of SRs dissected from females mated to either *iab-6*^{cocu} or control males. Significantly less SP is present in the SR of mates of *iab-6*^{cocu} males at 2h ASM compared to mates of control males, while the amount of SP present in the remainder of the reproductive tract is comparable, though

slightly higher in mates of *iab-6*^{cocu} males (Figure 5.4D; Figure S9 [46]). These results suggest that *iab-6*^{cocu} males transfer normal amounts of SP but that this SP fails to enter the SR. The reduction/absence of stored SP at later time points (1-7 days ASM) is likely responsible for the reduction in egg laying and the increase in receptivity seen in mates of *iab-6*^{cocu} males.

Sperm competition in mates of iab-6^{cocu} males.

SP also plays a role in sperm competition [60], which occurs when ejaculates from two males are present within the same female reproductive tract [61]. For example, in circumstances where SP null males are the first male to mate with a given female, they sire a higher percentage of the total progeny (P1, #progeny from first male / # total progeny) than control males [33]. To test whether the *iab-6*^{cocu} mutation also affects P1, we mated *iab-6*^{cocu} and control males to *cn bw* females and, after 3 days, allowed them to mate with *cn bw* males. The *iab-6*^{cocu} males had significantly higher P1 than control males consistent with a problem in SP presence or storage (Figure 5.6A).

Sperm storage in mates of iab-6^{cocu} males.

One possible explanation for the reduction in stored SP in mates of *iab-6*^{cocu} males is a defect in sperm entry into storage or an increase in the rate at which sperm are released from storage. To distinguish between these options, we counted sperm present in both female sperm storage organs at 2 hours, 4 days, and 10 days ASM. Females mated to either control or *iab-6*^{cocu} males store sperm at comparable levels at 2 hours ASM and appear to retain normal numbers through 4 days ASM (Fig.6B&C). These results suggest that initial sperm storage and release is normal in mates of *iab-6*^{cocu} males and that the reduced level of SP in the seminal receptacle at 2

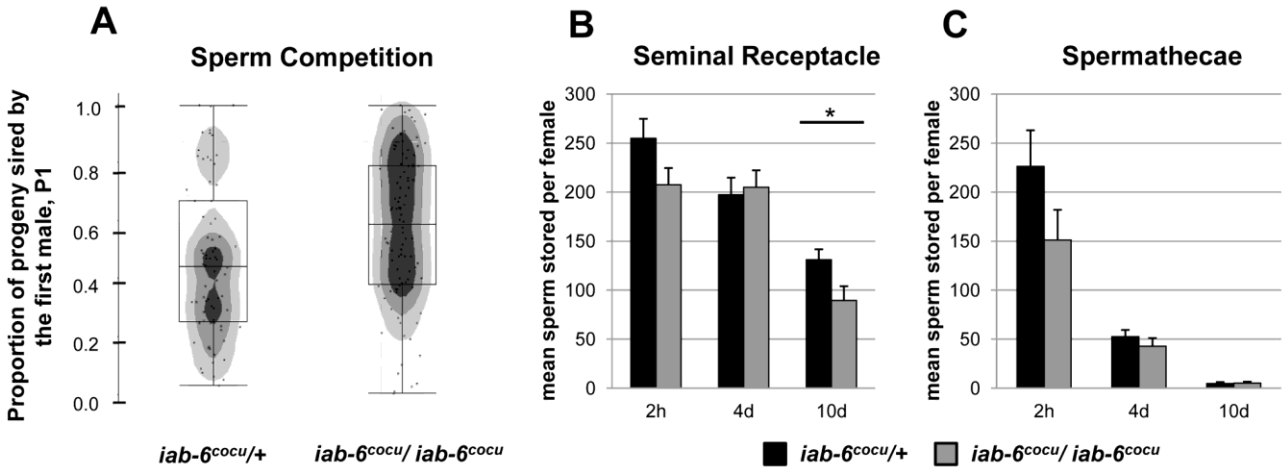


Figure 5.6: Sperm storage and use by mates of of *iab-6^{cocu}* or control males

A) For sperm competition assays *cn bw* females were first mated to either control (left) or *iab-6^{cocu}* (right) as the first male and allowed to mate a second time with a *cn bw* male. The proportion of progeny sired by *iab-6^{cocu}* males when acting as the first male (P1, # progeny from first male / total progeny) was significantly higher when compared to females who first mated with control males (WRST p= 0.038*, control N=74, *iab-6^{cocu}* N=98). **B&C)** Counts of sperm stored in mates of control (black) and *iab-6^{cocu}* (grey) males at 2h, 4d, and 10d ASM. **B)** Mates of *iab-6^{cocu}* males have wild type numbers of sperm present in the seminal receptacle at 2h (WRST p=0.10, control N=8, *iab-6^{cocu}* N=11) and 4d ASM (WRST p=0.96, control N=10, *iab-6^{cocu}* N=8) but fewer at 10d ASM (WRST p=0.017*, control N=19, *iab-6^{cocu}* N=12) when compared to mates of control males. **C)** Mates of *iab-6^{cocu}* males show wild type numbers of sperm stored in the spermathecae at all time points. 2h (WRST p=0.13, control N=7, *iab-6^{cocu}* N=10); 4d (WRST p=0.38, control N=10, *iab-6^{cocu}* N=7); 10d (WRST p=0.77, control N=17, *iab-6^{cocu}* N=16).

hours ASM is not due to a failure to adequately store sperm. Mates of *iab-6*^{cocu} males do not show the stereotypical sperm over-retention phenotype seen with knockdown of other LTR related proteins [33] and instead show a slight but significant decrease in stored sperm within the SR at 10 days ASM when compared to controls (Figure 5.6B). This is not wholly surprising, as the *iab-6*^{cocu} mutation does not result in the absence of a single gene product, but likely several that contribute to different aspects of the PMR. It is possible that the secondary cells produce, modify, or transfer some product necessary for sperm to be retained within the female sperm storage organs. When SP is absent, this product may be regulated improperly resulting in the typical sperm over-retention phenotype. A loss of, or reduction in this product, combined with the SP retention defect, may explain these results.

Several Acps display defects in glycosylation, stability, or protein abundance within the reproductive tract of females mated to iab-6^{cocu} males.

Our Western blots and RNAi data showed that three of the known LTR specific proteins (CG1656, CG1652, and CG17575) are produced in the secondary cells and that one (CG9997) is likely produced by main cells (Figure 5.4B&S7). We considered the possibility that failure to transfer one or all of these Acps to the female during mating could contribute to the SP storage defect seen in mates of *iab-6*^{cocu} males. To test this, we performed Western blots on the reproductive tracts of females mated to either *iab-6*^{cocu} or control males at 15m, 30m, and 1h ASM using antibodies to CG9997, CG1656, CG1652, and CG17575. Although all four Acps are transferred to females and are present in the female reproductive tract at all time points tested, their abundance, gel mobility, or processing appear to be abnormal in mates of *iab-6*^{cocu} males (Figure 5.7).

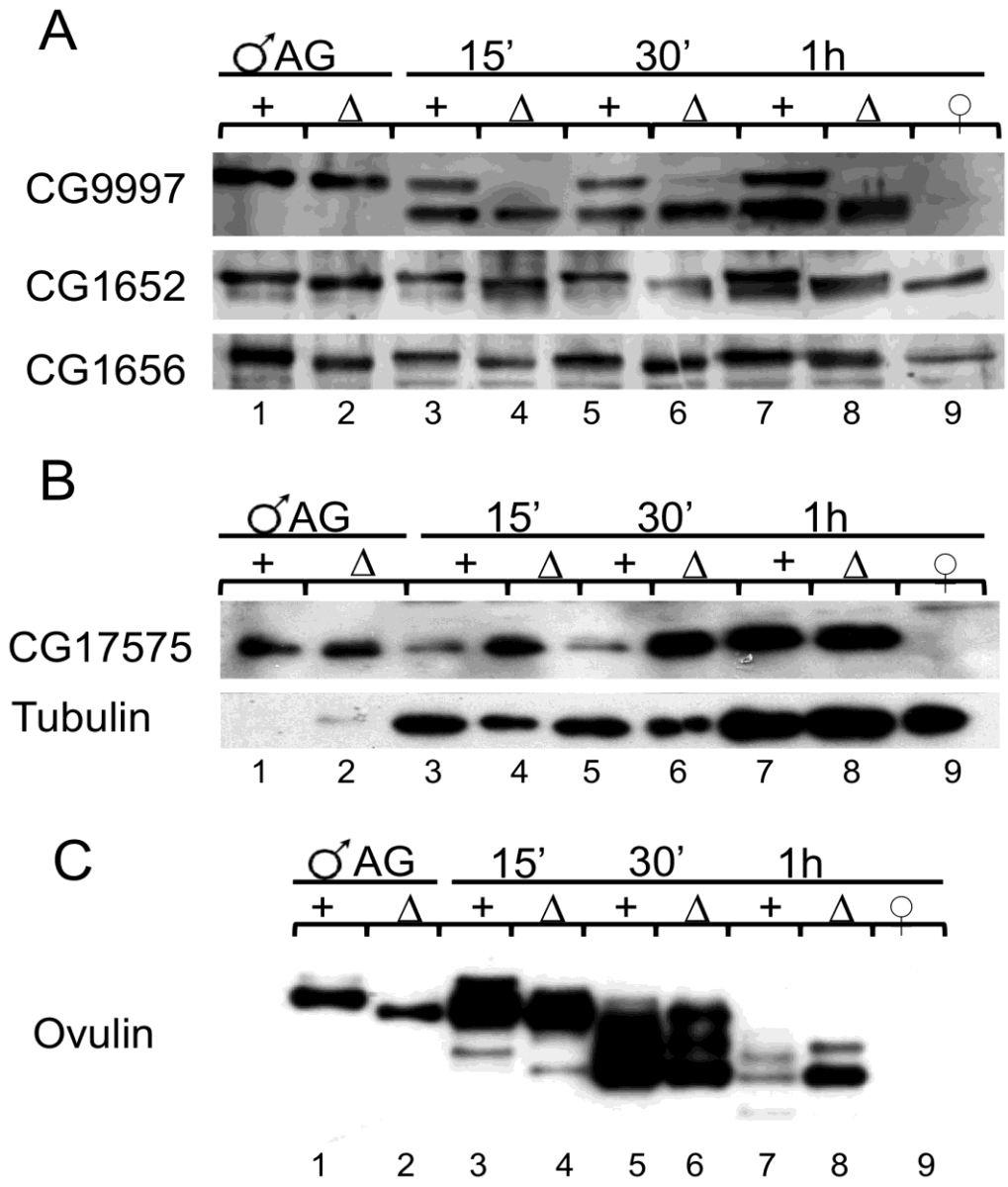
iab-6^{cocu} males carry out abnormal glycosylation of several Acps.

Both CG1656 and CG1652 run at a lower apparent molecular weight in *iab-6^{cocu}* males compared to control males (Figure 5.7A, Figure 5.4B). Ovulin, likewise, shows reduced apparent molecular weight in *iab-6^{cocu}* males (Figure 5.7C, Figure 5.4A). The gel mobility differences for these proteins in *iab-6^{cocu}* versus control males is evident both within AG extracts as well as across time points (15m, 30m, and 1h ASM) within the female reproductive tract of their mates. Ovulin is normally processed inside the female reproductive tract [62]. This processing appears to occur properly in mates of *iab-6^{cocu}* males, although the apparent molecular weight of some cleavage products is altered. It is unlikely that these differences in apparent molecular weight are caused by sequence differences or background effects. The controls used for these experiments are heterozygous for all of the chromosomes in the *iab-6^{cocu}* mutant line. Thus, if the gel mobility differences were caused by sequence differences, we would expect to see two bands indicating the WT and altered version of each protein. Further, ovulin and CG1656/1652 are located on separate arms of chromosome 2 and are necessary for different aspects of the PMR. Together, these observations suggest that the gel mobility differences may be a result of posttranslational modification.

Ovulin is a glycoprotein [54], but little is known about the posttranslational modifications of CG1656 and CG1652. To test whether differences in glycosylation underlie the gel mobility differences observed, we treated extracts from control and *iab-6^{cocu}* males with PNGaseF. This treatment, which removes N-linked glycosylation [63,64], resulted in loss of the apparent molecular weight differences between *iab-6^{cocu}* and control flies for all three proteins (Figure 5.8, CG1652 not shown). These results suggest that the secondary cells contribute to the regulation

Figure 5.7: Post translational modification, stability, and abundance of seminal fluid proteins in mates of *iab-6^{cocu}* or control males

Western blots using antibodies of known LTR-associated Acps CG9997, CG1656, CG1652, and CG17575 as well as STR Acp ovulin (Acp26Aa). Accessory gland extracts from a single control (lane 1) and *iab-6^{cocu}* male (lane 2) were used as positive controls and reproductive tract extracts from 4 virgin females (lane 9) were used as a negative control. Extracts from the reproductive tracts of females mated to control (+) or *iab-6^{cocu}* (Δ) were collected at 15' (lanes 3-4, 2 RTs per), 30' (lane 5-6s, 3 RTs per), and 1h ASM (lanes 7-8, 6 RTs per) . **A)** Full length CG9997 is produced by *iab-6^{cocu}* males but is not present in the reproductive tracts of their mates. The smaller processed form of CG9997 is present in mates of *iab-6^{cocu}* suggesting that CG9997 is transferred. Both CG1656 and CG1652 are transferred to females normally by *iab-6^{cocu}* males, but both of these proteins run at a lower apparent molecular weight than in control males. **B)** *iab-6^{cocu}* males transfer more CG17575 to their mates than control males. Tubulin was used as a loading control for the female reproductive tracts. **C)** Both mates of control and *iab-6^{cocu}* males receive ovulin. However, the ovulin produced by *iab-6^{cocu}* males also runs at a lower apparent molecular weight than in controls.


A
♂AG
15’
30’
1h
+ Δ + Δ + Δ + Δ ♀
CG9997
CG1652
CG1656
1 2 3 4 5 6 7 8 9
B
♂AG
15’
30’
1h
+ Δ + Δ + Δ + Δ ♀
CG17575
Tubulin
1 2 3 4 5 6 7 8 9
C
♂AG
15’
30’
1h
+ Δ + Δ + Δ + Δ ♀
Ovulin
1 2 3 4 5 6 7 8 9

of posttranslational modifications, and more specifically glycosylation, of Acps. They are also the first evidence of the presence of N-linked glycosylation on CG1656 and CG1652.

To verify that the glycosylation differences are caused specifically by the loss of Abd-B expression in the secondary cells, we also looked at the gel mobility of CG1656 and CG1652 in BAC rescued *iab-6*^{cocu} flies (Supplementary Figure S5, CG1652 not shown). The rescue BAC was able to restore proper glycosylation in the AGs of some males but not others. This is consistent with the partial rescues we have observed, especially if the glycosylation phenotype is dose dependant and the BAC males are on the threshold. Still, these results support the connection between Abd-B expression in the secondary cells and proper glycosylation of Acps and suggests that there may be a connection between these glycosylation differences and the PMR.

Mates of iab-6^{cocu} males display abnormal stability or abundance of some Acps.

In wild type males, CG9997 is transferred to females as a full length protein (45kD) and is processed in the female reproductive tract to a smaller form (36kD). Both products persist in the female for longer than 1h. Males with the *iab-6*^{cocu} mutation produce full length CG9997 but the full length form does not persist inside the female reproductive tract (Figure 5.7A). A similar increase in processing or instability of the full length product was seen in males that do not produce or transfer CG1656/CG1652 [39]; its biological relevance is currently unknown. Since *iab-6*^{cocu} males produce and transfer CG1656/CG1652, our results suggest that in addition to these two proteins, the products of the secondary cells are essential for regulating the stability of CG9997 inside the female. As observed with the other LTR Acps we assayed, CG17575 is produced and transferred to females by *iab-6*^{cocu} males. However, CG17575 is present in higher

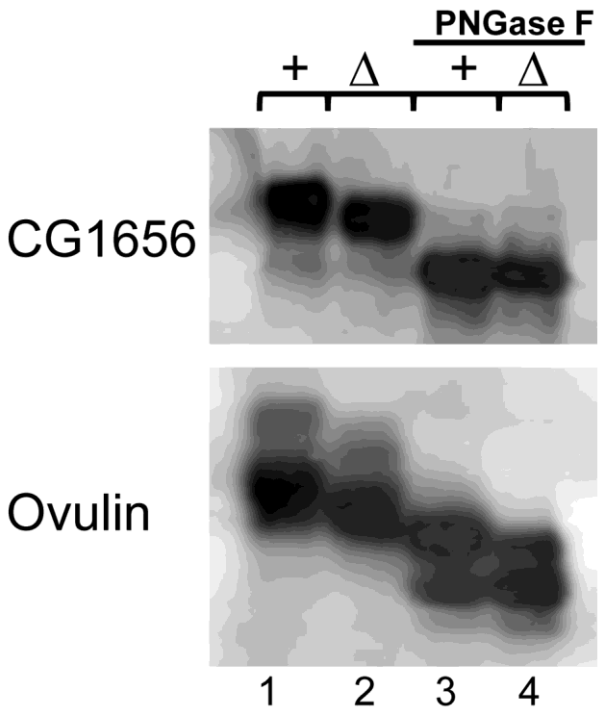


Figure 5.8: Glycosylation measurements of seminal fluid proteins in *iab-6*^{cocu} or control males.

PNGase F assays were used to examine glycosylation states of CG1656 and ovulin or determine whether N-linked glycosylation differences underlie the gel differences seen between control (+) and *iab-6*^{cocu} (Δ) males. Untreated (lanes 1-2) and PNGase F treated (lanes 3-4). In both cases the gel mobility differences seen between control and *iab-6*^{cocu} males are absent after PNGase F treatment suggesting that in *iab-6*^{cocu} males both proteins are improperly glycosylated.

amounts in the reproductive tract of females mated to *iab-6*^{cocu} males at all time points when compared to controls (Figure 5.7B; Figure S9). Whether this difference is due to increased transfer or failure to degrade CG17575 within the tract is unclear.

5.3 DISCUSSION

Here, we report that *Abd-B* is expressed in the secondary cells of the *Drosophila* male AG and, using mutations that specifically remove *Abd-B* from these cells, uncover roles for this previously unstudied but important reproductive cell type. Furthermore, we show that *Abd-B* expression in these mesodermally-derived cells does not fit the "initiator" paradigm developed for the segment-identity function of Hox genes in ectodermal tissues. And finally, we demonstrate that the secondary cells of the male AG synthesize products necessary for maintenance of the seminal fluid's effects on the female PMR.

New insights into Abd-B gene regulation

Due to the large size and complexity of the *Abd-B* regulatory region, we created a BAC-reporter construct to monitor *Abd-B* expression in the adult fly. When combined with fluorescent markers, this method allowed us to bypass the technical issues of antibody penetration and the laborious dissections needed for *in situ* hybridization or immunohistochemistry to identify a novel area of *Abd-B* expression in the adult. Overall, the BAC reporter is able to accurately reproduce the known, complex *Abd-B* expression pattern; indeed, our BAC construct seems to more-faithfully reproduce *Abd-Bm* expression than even a previously isolated transposon insert in the *Abd-B* promoter (*Abd-B-Gal4*^{LDN}) [65]. Furthermore, by combining our BAC reporter with

pre-existing deletion mutations, we were able to discover the function of a vital gene in an adult tissue without the need to create mitotic clones.

From the standpoint of Hox gene regulation, our discovery of the secondary cell enhancer is quite interesting because, unlike other cell-type specific enhancers from the BX-C, the secondary cell enhancer does not seem to be regulated by a domain initiator [6,41,51]. Most cell-type specific enhancers from the BX-C are not intrinsically restricted along the A-P axis. They are restricted only to a specific cell-type and gain A-P restriction through clustering in a BX-C domain. For example, in a transgene assay, an *Abd-B* enhancer from the *iab-7* domain (called 11X) drives expression in the tracheal placodes in all segments. However, in the BX-C, this enhancer seems to be active only in the *Abd-B* expression domain [6]. The clustering of enhancers to one area of the chromosome is thought to allow all of the enhancers to be coordinately regulated along the A-P axis as a domain through the changing of the local chromatin environment. The *Polycomb (Pc)* repression machinery is thought to be critical for this process by creating repressive chromatin over inactive domains ([6] and refs. therein). Specialized elements, called initiators, seem to read an A-P segmental address and act as domain activators, probably by preventing *Pc* repressive complexes from establishing on active domains [41].

The domain model predicts that deletion of an initiator element should prevent domain activation, leaving all enhancers in its domain inactive [41]. Based on this paradigm and the fact that the secondary cell enhancer was found in the *iab-6* domain, we expected that the deletion of the *iab-6* initiator would abolish *Abd-B* expression in the secondary cells. However, we found that *Abd-B* expression in the secondary cells of initiator mutants was normal. This finding argues against the strictest interpretation of the initiator model. We can propose several hypotheses to

resolve this discrepancy. For example, the *Pc* repression system is known to act on many genes during development, but its main targets appear to be the homeotic genes during the establishment of segment identity. It is possible that late in development, after cells have made initial cell fate decisions (and the segment identity role of the homeotic genes might be less important), the targets of *Pc* silencing might change. Such loosening of *Pc* silencing over the *Abd-B cis*-regulatory regions would allow previously silenced enhancers to become available for regulating *Abd-B* expression so that it could perform other functions, such as in the secondary cells.

Alternatively, the difference in *Abd-B* gene regulation that we observe in secondary cells may reflect the cellular origin of the secondary cells. Most BX-C *cis*-regulatory mutations were isolated based on cuticular phenotypes and confirmed using antibody staining in the epidermis and CNS. These tissues are of ectodermal origin, unlike the mesodermally-derived secondary cells [66]. Perhaps, the rules governing the coordination of Hox expression in the ectoderm do not hold true in the other germ layers. Consistent with this, BX-C genes are expressed differently in the gut visceral mesoderm than in the ectoderm [67].

Evolutionary considerations may provide some explanation for why the fly uses different means of controlling *Abd-B* expression in embryonic segment identity specification *vs*. in later reproductive tissues. *Abd-B* class Hox proteins play roles in the formation of the external genitalia in both arthropods and mammals [68,69,70,71,72]**.** Due to the similarity in expression pattern and function, it has been proposed that *Abd-B*'s role in the formation of the genitalia predates its role in segmental identity [69] [73]. Here, we have shown that *Abd-B* is also important for correct development of cells within the *Drosophila* male AG that produces many seminal fluid proteins required for male reproductive success. The mammalian orthologues of

Abd-B, the Hox9 to 13 class of genes, are expressed in the developing seminal vesicle and prostate gland, both seminal protein secreting organs [74,75]. The analogy in function between these organs, and their similarity in gene expression patterns suggests that the role of *Abd-B* class genes in the male reproductive tract might be an ancient, conserved function, potentially independent of its role in segmental identity. In this light, it would not be surprising that *Abd-B* regulation in the secondary cells escapes the domain regulation seen for *Abd-B* function in segment identity determination. The *cis*-regulatory domains for segment identity could have been added separately, possibly through co-opting the *abd-A* gene regulatory regions, as suggested by transvection studies [76,77]. In any case, the adding of *cis*-regulatory sequences and the consequent constraints of the domain model on *Abd-B* would necessarily have to preserve its late function in the secondary cells.

The function of the secondary cells of the male accessory gland

Previous studies have shown that the male AG produces Acps required to initiate and maintain a range of PMRs in females. Further, diptheria-toxin mediated-ablation of accessory gland main cells demonstrated that products of these cells are essential for the PMR. The importance of the main cells was further strengthened by the discovery that they produce the Acp sex peptide (SP), which is essential for many aspects of the PMR and whose persistence allows maintenance of PMR effects (i.e. the LTR) [29]. Additional Acps important for other aspects of the PMR (e.g. Acp36DE, ovulin) were also found to be produced by main cells. Thus, until now, the role of the secondary cells was unknown and no methods to directly target these cells were available.

Using the *iab-6*^{cocu} regulatory mutant of *Abd-B*, we demonstrated that the secondary cells make a unique contribution to maintenance of the female's post-mating changes in egg-laying, receptivity, sperm competition, and sperm storage (summarized in Figure 5.9). Our results are consistent with findings about secondary cell function obtained independently by Minami et al., from their study of *dve* mutants [78]. The inability of *iab-6*^{cocu} males to maintain the PMR in their mates is consistent with perturbation of the function of the "LTR network". These LTR network proteins are needed to promote the association of SP with sperm in the SR, an association that is required for SP-mediated maintenance of the PMR [22,39]. The results of our study also show that three of these LTR network Acps, CG1656, CG1652, and CG17575 are all produced in the secondary cells while CG9997 and Seminase are primarily or exclusively main cell expressed. This is the first direct evidence that Acps from both cell types work together in a complex pathway. While all reported LTR specific network proteins (CG9997, CG1656, CG1652, and CG17575) are present in the *iab-6*^{cocu} mutant, they are all abnormal, either in amount/stability inside the female reproductive tract or in glycosylation state; how these changes result in the SP storage defect is an important area for future study.

The phenotype of *iab-6*^{cocu} males shows one exception to the standard LTR phenotype. Mates of SP null males (or males knocked down for any of the other 4 LTR related proteins) have high rates of sperm retention; this phenotype is lacking in mates of *iab-6*^{cocu} mutants. The mechanisms behind the release of sperm from storage and this sperm retention phenotype are unknown. Loss of *Abd-B* expression most likely impacts a wide variety of Acps and potentially cellular functions associated with vacuoles (such as transmembrane transport). Thus, it is possible that one or a combination of the proteins affected by this mutation (some as yet unidentified) may negatively impact normal sperm storage independent of the influence of *Abd-B*

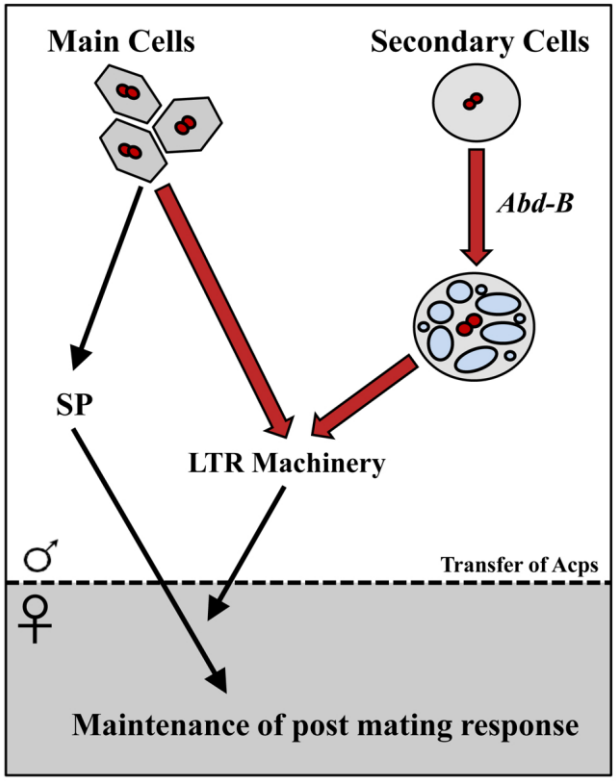


Figure 5.9: Chapter 5 - Summary/model.

Bold arrows delineate the new findings in this paper. The Drosophila male accessory gland consists of two secretory cell types: main cells and secondary cells. Main cells produce seminal proteins essential for inducing post-mating responses (PMR) in mated females; the function (if any) of secondary cells was unknown. Sex Peptide (SP) from the main cells induces many aspects of PMR, but persistence of its effects requires other seminal proteins (the LTR machinery: CG9997, CG17575, CG1652 and CG1656), whose cellular source was unknown.

Here, we showed that the Hox gene *Abd-B* is essential for normal development of the secondary cells, but with regulatory characteristics (not shown in this figure) that differ from those used in segment identity. By deleting the secondary-cell regulatory element of *Abd-B*, we obtained males with abnormal secondary cells. Mates of those males failed to maintain the PMR, indicating that secondary cells play an essential role in reproduction: their products, along with main cell products, allow persistence of SP in mated females, thus prolonging their post-mating responses.

on the storage of SP in the seminal receptacle. This could result in masking the over-retention of sperm seen in SP nulls due to the loss or abnormal function of a protein or proteins necessary for the retention phenotype to occur. Further work investigating the role of individual secondary cell associated Acps in the LTR, as well as sperm storage, may be helpful in determining how these processes function.

The *iab-6*^{cocu} mutant revealed another unique role for secondary cells: regulating glycosylation of at least three seminal proteins (ovulin, CG1656, and CG1652); the impact of glycosylation on the functions of these proteins is as yet unknown. Our findings show that CG1656 and CG1652 are produced by the secondary cells. However, ovulin is produced in main cells as well as secondary cells, and is detected in the AG lumen as well as in the vacuoles of the secondary cells [54]. As such, we suggest three possibilities for how secondary cells might mediate glycosylation: 1) through the secretion of glycosylation substrates that can be taken up and used by main cells, 2) through the secretion of glycosyation regulators directly into the lumen where they could modify Acps from both cell types that are present there, or 3) that Acps like ovulin and other main cell products are taken up into these vacuoles and then modified before being secreted into the lumen as mature, glycosylated proteins. Future dissection of the glycosylation phenotypes in the *iab-6*^{cocu} mutant males will help shed light on the role glycosylation plays in regulating the PMR and how this process is regulated in the tissue as a whole.

It is important to note that our *Abd-B* regulatory mutant still has secondary cells, or at least precursor cells poised to become secondary cells. The initial differentiation step that allows for *Abd-B* expression in the secondary cells is not perturbed by the *iab-6*^{cocu} deletion. As such, although we have already found a new and unanticipated role for secondary cells in regulating

the PMR, it is possible that these cells play additional roles. Future cell ablation experiments using tools derived from this study will allow tests of such additional roles.

In conclusion, we have shown that each of the two cell types of the *Drosophila* AG plays important roles in producing the female PMR: the main cells producing several Acps to initiate and maintain the PMR, and the secondary cells providing products to aid in temporally extending the response. We expect that through the action of various combinations of transcription factors, like *Abd-B*, the two cell-type lineages have diverged into distinct, and specialized cell-types. Although these two cell types perform vitally intertwined functions, they are maintained as separate cell types. This suggests that there may be a requirement for compartmentalization of their functions or products, or that the two cell types evolved separately for some other purpose and were functionally associated afterward. It is, therefore, of great future interest to identify the specific products of each of these cell types and to determine how they work in conjunction to mediate the full reproductive effect of seminal secretions.

Acknowledgements

I thank H. Nakagoshi for communicating results prior to publication, for coordinating reports of their data with ours, and for comments on this manuscript. I thank F. Avila, G. Findlay, D. Rubinstein, C. Sartain, and A. Mutero for comments on the manuscript, and members of the Karch and Wolfner labs for helpful suggestions during the course of this research. I thank C. Iampietro for making the *iab-6*$^{\Delta 5}$ mutation . I also thank E. Favre and J. Faustino for excellent technical assistance. This work was supported by grants from the Donation Claraz, the State of Geneva and the Swiss National Fund for Research to FK, and by NIH grants R01HD-038921 to MFW and R01HD-059060 to MFW and A.G. Clark. For part of

this work, JLS was supported on a traineeship on National Institutes of Health training grant T32-GM07617.

5.4 MATERIALS AND METHODS

Creation of Abd-B rescue BAC.

The *Abd-B* rescue BAC was constructed from BACR24L18 (size - 171936bp; AC095018), which contains the *Abd-B* region of the BX-C. We first constructed a pW25-based vector [79] to be used to add sequences to the BAC for site-specific integration into the *Drosophila* genome. The original *white* gene carried by pW25 was replaced by a *Su(Hw)* insulator-flanked *white* gene from the SUPor-P plasmid [80] by amplifying it with *NotI-forward* and *AscI-PmeI reverse* (see Table 1 for primer list). These primers carry restriction sites for insertion of the product into pw25 after appropriate enzyme digestion. A fragment containing the kanamycin resistance gene (KanR) was amplified from pIGCG21 [81] using the following primers: PmeI5'Kan, 5'attB3'KanAS. Next, an AttB sequence was PCR amplified using the primers 3'Kan5'attBS, and PmeI3'attBAS. The KanR and AttB fragments were then mixed together with the PmeI5'Kan and PmeI3'attBAS primers for a final overlap PCR reaction. The resulting PCR fragment, containing the KanR gene and an AttB sequence flanked by PmeI sites, was cloned into pGemTeasy. After sequencing, this fragment was excised with PmeI and cloned into the unique PmeI site of the modified pW25 vector (resulting in pW25/Kan-AttB).

To mediate recombination in bacteria, two homology regions were then added to this vector. First, an 859bp fragment from the *iab4* region (*iab4* HR: 101409-102267 coordinates in the BACR24L18) was made by PCR with the primers: NotI *iab4* and EagI *iab4*. The resulting PCR

fragment was cut with NotI and inserted into the unique NotI site of the modified pW25. A second homology fragment was designed to target the SacBII gene present on the BAC backbone. Using the PCR primers: MluI SacBII and AscI SacBII, a 525bp fragment from SacBII was amplified (SacBII HR: 921-1445 coordinates in the BACR24L18). The resulting 525bp fragment was digested with AscI and MluI, and inserted into the unique AscI site of pW25/Kan-AttB. Clones for both homology regions were selected in an orientation required by the homologous recombination process to function correctly. The completed construct was digested with the ISce-I endonuclease and the fragment containing the homology regions flanking the KanR gene, the white gene and the AttB site was gel purified. This fragment was then used to recombineer BAC24L18 using the protocol of Soren Warming [82]. The resulting BAC, called *iab4-SacBII BACR24L18D* (108528bp), contains the region of the BX-C from about iab-4 to the Abd-B m promoter. The overall structure of this BAC was verified by restriction enzyme mapping [using three restriction digests (EcoRI, XmaI, BamHI) (data not shown)].

Abd-B Gal4 reporter BAC.

Using *iab4-SacBII BACR24L18D* as a base we used recombineering to replace the start codon of the *Abd-Bm* isoform with sequence encoding the Gal4 transcription factor. First, a negative/positive selection cassette was created using the SacB gene and the Ampicillin resistance (AmpR) gene. SacBII was digested out of the plasmid pSK2-SACBK MAR using BamHI and EcoRI and cloned in pHSS7 [83]. The AmpR gene was then amplified with primers (*Amp BamHI* and *Amp XmaI)* carrying a BamHI and XmaI site. The amplified AmpR fragment was digested with BamHI and XmaI and cloned into a pHSS7-SacBII digested with the same enzymes, producing pHSS7-SacBII/Amp.

The SacBII/Amp cassette was flanked by two large homology regions using an overlap PCR strategy, as follows: The primers NHR L frw NotI/long and NHR L rev OL/long were first used to amplify the region 39201bp to 40325bp of iab4-SacBII BACR24L18D. At the same time, the primers NHR R frw OL/long and NHR R rev XmaI/long were used to amplify the 40326 to 41590bp region of *iab4*-SacBII BACR24L18D. The two reaction products contain a 27bp region of complementarity to mediate overlap PCR. Thus, the two fragments were mixed together with the primers NHR L frw NotI/long and NHR R rev XmaI/long in an overlap PCR reaction. The resulting fragment, containing the two homology domains fused together, was digested with NotI and XmaI and cloned into pHSS7 (pHSS7-NHR-L/NHR-R). Next, from the previously created pHSS7-SacBII/Amp, the double selection cassette was amplified with primers *NSx/A new F EcoRI* and *NCSx/A new R XbaI*. Digesting this PCR fragment with EcoRI and XbaI produced a fragment that could be inserted between the two homology domains of pHSS7-NHR-L/NHR-R creating pHSS7-NHR-L/SacBII-Amp/NHR-R.

The primers *F300 NS/A rec* and *R300 NS/A rec* were used to amplify a fragment from pHSS7-NHR-L/SacBII-Amp/NHR-R for recombineering. This fragment contained 300bp of the homology regions on both sides of the cassette carrying the SacBII and AmpR genes. Once again recombineering was performed using the protocol of Soren Warming [82], where the target BAC was *iab4-SacBII BACR24L18D*. BAC DNA purified from the resulting ampicillin resistant/sucrose sensitive colonies were verified by extensive restriction enzyme digests. The new BAC was named *iab4-SacBII BACR24L18D N S/A ins*.

To replace the SacBII/AmpR cassette, the Gal4 gene (with a synthetic polyA tail) was PCR amplified from the plasmid pTnT Gal4 (unpublished, pTNT base vector from *Promega Corp.*, Madison, Wisconsin, USA) using the following primers: HR-R Gal4rep and HR-L Gal4rep. These primers contain 55bp homology regions to mediate recombineering to the BX-C sequences just flanking the SacBII/AmpR cassette. The resulting targeting fragment was then phosphorylated by T4 kinase in order to improve the recombination reactions. After standard preparation of the recombineering DY380 strain containing *iab4-SacBII BACR24L18D N S/A ins*, bacterial colonies were selected on LB agar plates containing 10% sucrose. Restriction digestion of the candidate colonies with BamHI was performed in order to confirm the correct replacement of the SacBII/Amp cassette with Gal4.

Injections of BACs.

Using the PhiC31 system ([45]; www.flyc31.org), site 51C on the second chromosome was chosen for integration of the BACs into the fly genome. For better integration frequencies, all BACs were isolated on the day of injection using the NucleoBond PC 20 (Macherey-Nagel ref 740571) miniprep kit and resuspended in injection buffer [84]. Embryos were injected with BAC DNA (at about 50-100ng/ul) through the chorion using the Eppendorf system (FemtoJet & TransferMan NK 2) equipped with Femtotips II glass needles. Integration efficiency was about 5%, based on the total number of fertile adults that yielded at least one integrant.

Creation of a specific secondary cell Gal4 driver based on the cocu enhancer.

The 2.8kb putative enhancer sequence removed in *iab-6*^{cocu} was amplified by PCR using primers D5 F and D5 R. Both of these primers contain a BamHI site at their 5' ends. The amplified DNA

fragment (called D5) was cloned into the BamHI site of the pChs-Gal4 plasmid, which contains a minimal Hsp70 promoter upstream of the coding sequence for Gal4 and the HSP70 3'UTR (Drosophila Genomics Resource Center [85]. Although clones with the enhancer in both orientations were isolated, we proceeded using a clone where the D5 R primer containing end was closest to Gal4. The D5-Gal4 cassette was then digested out of the pChs-Gal4 vector with NotI and cloned into the NotI site of pattB [45]. An insertion with the Gal4 coding sequence next to the *white* gene was selected for injection. This construct was integrated by Genetic Services Inc (Cambridge, Mass) into the VK00001 (59D3) platform [86]. The resulting integrant is named D5rsG4rs and referred as to D5-Gal4 in the text..

Fly crosses and strains.

All crosses were done using standard genetic techniques. *iab-7*^{Sz}, *iab-6,7*^{IH}, *iab-5,6*^{J82}, and *iab-4,5,6*^{DB} are described in [6]. The lines *iab-6*$^{\Delta 5}$ and *iab-6*^{4} are described in [41]. The line *iab-6*$^{\Delta 5}$ was described as a deficiency without any phenotypic consequence. Following the LTR phenotype identified in this work the line was renamed *iab-6*^{cocu} (reflecting that mates of these males fail to reject other suitors; "cocu" means "cuckold" in French). The BAC-*AbdB*$^{\text{Gal4}}$,w^{+}, UAS-GFP/Cy line carrying the *Abd-B* Gal4 BAC reporter and a UAS-GFP marker on the second chromosome was created for this study by recombining a chromosome carrying the BAC and a UAS-GFP chromosome. The BAC reporter chromosome cannot exist as a homozygote. The 4.4E transgenic *lacZ* reporter line is described in [6]. The Gal4 expressing lines driven by a *paired* enhancer ((w^{-}; prd-mf5.2,w^{+}/CyO), (w^{-};prd-mf5.4,w^{+}), (w^{-};prd-mf5.5,w^{+}), (w^{-};prd-mf9.3,w^{+}), (w^{-};prd-mf9.7,w^{+})) were obtained from Makus Noll's laboratory [48]. They were used in a cross

with a UAS-*AbdBm* [87] flies for the experiment in which we tested for the ability of *Abd-Bm* to transform main cells into secondary cells.

All flies for fertility and fecundity assays, tests of receptivity and sperm competition, Western blotting, sperm counts, and PNGase F assays were raised at room temperature (23±1°C). Females were aged 3-5 days from eclosion in groups of 7-11 in glass vials on standard yeast-glucose media with added yeast. Males were aged 3-5 days from eclosion in groups of 10-20 in glass vials on standard yeast-glucose media.

Antibody, X-Gal and** FM4-64 **staining.

Antibody and X-Gal staining on embryos and dissected accessory glands was performed as described in [12] and [42] respectively, using a 20min fixation. The *Abd-B* primary antibody, obtained from the Developmental Studies Hybridoma Bank, was diluted 1:4. Goat-anti-mouse secondary antibody, coupled to Alexa Fluor 488/555 X (Invitrogen AG), used to reveal *Abd-B* localization was used at 1:500 dilution. Goat HRP coupled anti-mouse was obtained from Biorad and used at 1:1'500 dilution. Staining with FM4-64 dye was done by placing a drop of the dye onto a microscope slide and placing a freshly dissected gland into it. The glands were immediately covered with a cover slip and visualized using fluorescent microscope at 555nm.

Fertility/Fecundity, Receptivity, Sperm Counts, Western Blotting, and Sperm Competition Assays:

In all assays, we used 3-5 day old virgin females from a wild type strain (Canton-S for fertility/fecundity assays, receptivity assays, sperm counts, and for Western blotting experiments; and *cn bw* for sperm competition assays). Females were placed singly in glass vials with food and allowed access to an *iab-6*^{cocu}, control male (heterozygous for the *iab-6*^{cocu} mutation), DTA-E, or WT Canton-S male. Pairs were watched to confirm that mating had occurred. The male was removed upon dismounting. All statistical analysis was performed with the Jmp9 software. In the fertility/fecundity assays, after mating, individual females were housed for 24 hours in glass vials on yeast-glucose media. After 24 hours each female was transferred to a fresh vial, and the eggs laid in the previous vial were counted. This process was repeated for a total of 10 days. Upon eclosion, all progeny from each vial were counted. Hatchability (# progeny / # eggs) was calculated per day and across the 10-day period for each female. Values greater than 1 represent instances where the number of progeny produced exceeded the number of eggs observed. This is an accurate representation of counter error, and was not normalized to 1. Small levels of counter error has a greater impact on hatchability for females that lay few eggs. Comparisons of egg and progeny production between control and experimental females were performed using a Wilcoxon non-parametric test and statistics comparing the overall 10 day trends were performed using a repeated measures ANOVA.

Receptivity assays*:*

After mating, individual females were kept in a vial on yeast-glucose media for 1 day, 4 days, or 10 days after the start of mating (ASM). Each female was then moved to a fresh vial and provided with a single Canton S male. After addition of the single Canton S male, couples were observed at 15 min time intervals for one hour, and the proportion of successful matings was

recorded. The original vials were kept to check for progeny from the first mating. Females that did not produce viable progeny were discarded from the assay. Comparisons between the remating frequency of control and experimental females were conducted using a one way ANOVA.

Sperm counts.

After mating females were either frozen in liquid nitrogen at 2h ASM or kept in glass vials on yeast-glucose media for 4 days and 10 days ASM and then frozen. The female reproductive tracts were removed and stained with orcein (as described in [33,88,89]. Sperm were visualized and counted using a transillumination microscope at 1000x magnification. Comparisons between the number of sperm present in control and experimental females were performed using a Wilcoxon non-parametric tests.

Sperm competition assays:

After mating, individual females were housed for 3 days on yeast-glucose media. Females were then allowed access to a single *cn bw* male for 7 hours. Couples were observed for the first 4 hours in 15 min time intervals to determine the percent remating. After 7 hours the *cn bw* male was removed and the females were transferred individually to fresh vials and allowed to lay eggs for 4 days. They were then transferred individually to fresh food vials and allowed to lay eggs for an additional 4 days. Progeny were collected from each vial and assessed for the presence of red eyes (control or *iab-6*^{cocu} male sire) or white eyes (*cn bw* sire). P1 was calculated as # progeny sired by the first male / total progeny. Comparisons for P1 were performed using a one way ANOVA and by Wilcoxon non-parametric tests.

Western blots:

Females were frozen in liquid nitrogen at 15', 30', 1h, and 2h time points ASM and stored at -20°C until dissection. For later time points (1-7 days), females were kept individually on yeast-glucose media at room temperature before being frozen in liquid nitrogen and stored at -20°C until dissection. Males were also frozen in liquid nitrogen and stored at -20°C until dissection. Preparation of protein samples and Western blot analyses were performed as in [38,39] except that gels in this study were 5-15% acrylamide gradient gels and were run at 100v for 1.5-2 hours. Due to size differences in organs across males and the lack of an optimal loading control for male accessory glands, all comparative samples contain an identical number of reproductive tracts (male or female).

PNGase F assays were performed using reagents from New England Biolabs Inc. Male accessory glands from 10 *iab-6*^{cocu} males and 10 *iab-6*^{cocu} heterozygous control males were dissected in 1xPBS and transferred to 10ul 1x Glycoprotein Denaturing Buffer (GDB). Samples were ground and heated in GDB for 10min at 100°C. Then 2ul 10xG7 Reaction Buffer, 2ul 10% NP40, and 4ul ddH_2O were added. Each sample was split into 9ul aliquots. 1ul PNGase F was added to one aliquot and 1ul ddH_2O was added to the other. All samples were then incubated for 1 hour at 37°C and frozen at -80°C overnight. 10ul SDS sample buffer was added to each sample and then the samples were boiled for 5min at 100°C. Western blots were performed as previously described except that 10.6% acrylamide gels were used and run at 40v for 16h to ensure adequate separation.

4.5 REFERENCES

1. Sanchez-Herrero E, Vernos I, Marco R, Morata G (1985) Genetic organization of Drosophila bithorax complex. Nature 313: 108-113.

2. Karch F, Weiffenbach B, Peifer M, Bender W, Duncan I, et al. (1985) The abdominal region of the bithorax complex. Cell 43: 81-96.

3. Lewis EB (1978) A gene complex controlling segmentation in Drosophila. Nature 276: 565-570.

4. Sanchez-Herrero E (1991) Control of the expression of the bithorax complex genes abdominal-A and abdominal-B by cis-regulatory regions in Drosophila embryos. Development 111: 437-449.

5. Celniker SE, Sharma S, Keelan DJ, Lewis EB (1990) The molecular genetics of the bithorax complex of Drosophila: cis- regulation in the Abdominal-B domain. Embo J 9: 4277-4286.

6. Mihaly J, Barges S, Sipos L, Maeda R, Cleard F, et al. (2006) Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. Development 133: 2983-2993.

7. Maeda RK, Karch F (2006) The ABC of the BX-C: the bithorax complex explained. Development 133: 1413-1422.

8. Karch F, Galloni M, Sipos L, Gausz J, Gyurkovics H, et al. (1994) Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of Drosophila melanogaster. Nucleic Acids Res 22: 3138-3146 Issn: 0305-1048.

9. Gruzdeva N, Kyrchanova O, Parshikov A, Kullyev A, Georgiev P (2005) The Mcp Element from the bithorax Complex Contains an Insulator That Is Capable of Pairwise Interactions and Can Facilitate Enhancer-Promoter Communication. Mol Cell Biol 25: 3682-3689.

10. Busturia A, Bienz M (1993) Silencers in abdominal-B, a homeotic Drosophila gene. Embo J 12: 1415-1425 Issn: 0261-4189.

11. Hagstrom K, Muller M, Schedl P (1997) A *Polycomb* and GAGA dependent silencer adjoins the *Fab-7* boundary in the *Drosophila* bithorax comple. Genetics 146: 1365-1380.

12. Hagstrom K, Muller M, Schedl P (1996) *Fab-7* functions as a chromatin domain boundary to ensure proper segment specification by the *Drosophila* bithorax complex. Genes Dev 10: 3202-3215.

13. Zhou J, Ashe H, Burks C, Levine M (1999) Characterization of the transvection mediating region of the abdominal- B locus in Drosophila. Development 126: 3057-3065.

14. Zhou J, Barolo S, Szymanski P, Levine M (1996) The Fab-7 element of the bithorax complex attenuates enhancer-promoter interactions in the Drosophila embryo. Genes Dev 10: 3195-3201.

15. Gyurkovics H, Gausz J, Kummer J, Karch F (1990) A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. Embo J 9: 2579-2585 Issn: 0261-4189.

16. Barges S, Mihaly J, Galloni M, Hagstrom K, Müller M, et al. (2000) The *Fab-8* boundary defines the distal limit of the bithorax complex *iab-7* domain and insulates *iab-7* from initiation elements and a PRE in the adjacent *iab-8* domain. Development 127: 779-790.

17. Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF (2011) Insect seminal fluid proteins: identification and function. Annu Rev Entomol 56: 21-40.

18. Sirot LK, LaFlamme BA, Sitnik JL, Rubinstein CD, Avila FW, et al. (2009) Molecular social interactions: Drosophila melanogaster seminal fluid proteins as a case study. Adv Genet 68: 23-56.

19. Kalb JM, DiBenedetto AJ, Wolfner MF (1993) Probing the function of Drosophila melanogaster accessory glands by directed cell ablation. Proc Natl Acad Sci U S A 90: 8093-8097.

20. Liu H, Kubli E (2003) Sex-peptide is the molecular basis of the sperm effect in Drosophila melanogaster. Proc Natl Acad Sci U S A 100: 9929-9933.

21. Chapman T, Bangham J, Vinti G, Seifried B, Lung O, et al. (2003) The sex peptide of Drosophila melanogaster: female post-mating responses analyzed by using RNA interference. Proc Natl Acad Sci U S A 100: 9923-9928.

22. Peng J, Chen S, Busser S, Liu H, Honegger T, et al. (2005) Gradual release of sperm bound sex-peptide controls female postmating behavior in Drosophila. Curr Biol 15: 207-213.

23. Merle J (1968) [Ovarian function and sexual receptivity of Drosophila melanogaster after implantation of fragments of the male genital tract]. J Insect Physiol 14: 1159-1168.

24. Chen PS, Stumm-Zollinger E, Aigaki T, Balmer J, Bienz M, et al. (1988) A male accessory gland peptide that regulates reproductive behavior of female D. melanogaster. Cell 54: 291-298.

25. Garcia-Bellido A (1964) [the Secretion of Paragonia as a Stimulus for the Fecundity of Drosophila Melanogaster Females]. Z Naturforsch B 19: 491-495.

26. Xue L, Noll M (2000) Drosophila female sexual behavior induced by sterile males showing copulation complementation. Proc Natl Acad Sci U S A 97: 3272-3275.

27. Bertram MJ, Akerkar GA, Ard RL, Gonzalez C, Wolfner MF (1992) Cell type-specific gene expression in the Drosophila melanogaster male accessory gland. Mech Dev 38: 33-40.

28. DiBenedetto AJ, Harada HA, Wolfner MF (1990) Structure, cell-specific expression, and mating-induced regulation of a Drosophila melanogaster male accessory gland gene. Dev Biol 139: 134-148.

29. Styger D (1992) Molekulare analyse des Drosophila melanogaster sex-peptid Gens. Zürich, Switzerland: University of Zürich.

30. Heifetz Y, Vandenberg LN, Cohn HI, Wolfner MF (2005) Two cleavage products of the Drosophila accessory gland protein ovulin can independently induce ovulation. Proc Natl Acad Sci U S A 102: 743-748.

31. Chapman T, Herndon LA, Heifetz Y, Partridge L, Wolfner MF (2001) The Acp26Aa seminal fluid protein is a modulator of early egg hatchability in Drosophila melanogaster. Proc Biol Sci 268: 1647-1654.

32. Isaac RE, Li C, Leedale AE, Shirras AD (2010) Drosophila male sex peptide inhibits siesta sleep and promotes locomotor activity in the post-mated female. Proc Biol Sci 277: 65-70.

33. Avila FW, Ravi Ram K, Bloch Qazi MC, Wolfner MF (2010) Sex peptide is required for the efficient release of stored sperm in mated Drosophila females. Genetics 186: 595-600.

34. Rogina B (2009) The effect of sex peptide and calorie intake on fecundity in female Drosophila melanogaster. ScientificWorldJournal 9: 1178-1189.

35. Kubli E (1992) The sex-peptide. Bioessays 14: 779-784.

36. LaFlamme BA (2011 submitted) The Drosophila seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. Submitted.

37. Ram KR, Wolfner MF (2007) Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. Integrative and Comparative Biology 47: 427-445.

38. Ravi Ram K, Ji S, Wolfner MF (2005) Fates and targets of male accessory gland proteins in mated female Drosophila melanogaster. Insect Biochem Mol Biol 35: 1059-1071.

39. Ram KR, Wolfner MF (2009) A network of interactions among seminal proteins underlies the long-term postmating response in Drosophila. Proc Natl Acad Sci U S A 106: 15384-15389.

40. Ravi Ram K, Sirot LK, Wolfner MF (2006) Predicted seminal astacin-like protease is required for processing of reproductive proteins in Drosophila melanogaster. Proc Natl Acad Sci U S A 103: 18674-18679.

41. Iampietro C, Gummalla M, Mutero A, Karch F, Maeda RK (2010) Initiator elements function to determine the activity state of BX-C enhancers. PLoS Genet 6: e1001260.

42. Galloni M, Gyurkovics H, Schedl P, Karch F (1993) The bluetail transposon: evidence for independent cis-regulatory domains and domain boundaries in the bithorax complex. Embo J 12: 1087-1097 Issn: 0261-4189.

43. Zavortink M, Sakonju S (1989) The morphogenetic and regulatory functions of the Drosophila Abdominal- B gene are encoded in overlapping RNAs transcribed from separate promoters. Genes Dev 3: 1969-1981.

44. Foronda D, Estrada B, de Navas L, Sanchez-Herrero E (2006) Requirement of Abdominal-A and Abdominal-B in the developing genitalia of Drosophila breaks the posterior downregulation rule. Development 133: 117-127.

45. Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A 104: 3312-3317.

46. Gligorov D, Sitnik JL, Maeda RK, Wolfner MF, Karch F (2013) A novel function for the hox gene abd-B in the male accessory gland regulates the long-term female post-mating response in Drosophila. PLoS Genet 9: e1003395.

47. Martin CH, Mayeda CA, Davis CA, Ericsson CL, Knafels JD, et al. (1995) Complete sequence of the bithorax complex of Drosophila. Proc Natl Acad Sci U S A 92: 8398-8402 Issn: 0027-8424.

48. Jiao R, Daube M, Duan H, Zou Y, Frei E, et al. (2001) Headless flies generated by developmental pathway interference. Development 128: 3307-3319.

49. Peifer M, Karch F, Bender W (1987) The bithorax complex: control of segmental identity. Genes Dev 1: 891-898.

50. Bender W, Hudson A (2000) P element homing to the Drosophila bithorax complex. Development 127: 3981-3992.

51. Simon J, Peifer M, Bender W, O'Connor M (1990) Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis. Embo J 9: 3945-3956 Issn: 0261-4189.

52. Singh PB, Brown D (1997) Modelling the activity of the Ultrabithorax parasegment-specific regulatory domains around their anterior boundaries. J Theor Biol 186: 397-413.

53. Wolfner MF, Harada HA, Bertram MJ, Stelick TJ, Kraus KW, et al. (1997) New genes for male accessory gland proteins in Drosophila melanogaster. Insect Biochem Mol Biol 27: 825-834.

54. Monsma SA, Harada HA, Wolfner MF (1990) Synthesis of two Drosophila male accessory gland proteins and their fate after transfer to the female during mating. Dev Biol 142: 465-475.

55. Manning A (1967) The control of sexual receptivity in female Drosophila. Anim Behav 15: 239-250.

56. Yamamoto MT, Takemori N (2010) Proteome profiling reveals tissue-specific protein expression in the male reproductive system of Drosophila melanogaster. Fly (Austin) 4: 36-39.

57. Walker MJ, Rylett CM, Keen JN, Audsley N, Sajid M, et al. (2006) Proteomic identification of Drosophila melanogaster male accessory gland proteins, including a pro-cathepsin and a soluble gamma-glutamyl transpeptidase. Proteome Sci 4: 9.

58. Takemori N, Yamamoto MT (2009) Proteome mapping of the Drosophila melanogaster male reproductive system. Proteomics 9: 2484-2493.

59. Findlay GD, Yi X, Maccoss MJ, Swanson WJ (2008) Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. PLoS Biol 6: e178.

60. Chow CY, Wolfner MF, Clark AG (2010) The genetic basis for male x female interactions underlying variation in reproductive phenotypes of Drosophila. Genetics 186: 1355-1365.

61. Parker GA (1970) Sperm competition and its evolutionary consequences in the insects. Bio rev 45: 525-567.

62. Park M, Wolfner MF (1995) Male and female cooperate in the prohormone-like processing of a Drosophila melanogaster seminal fluid protein. Dev Biol 171: 694-702.

63. Maley F, Trimble RB, Tarentino AL, Plummer TH, Jr. (1989) Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. Anal Biochem 180: 195-204.

64. Plummer TH, Jr., Tarentino AL (1991) Purification of the oligosaccharide-cleaving enzymes of Flavobacterium meningosepticum. Glycobiology 1: 257-263.

65. de Navas L, Foronda D, Suzanne M, Sanchez-Herrero E (2006) A simple and efficient method to identify replacements of P-lacZ by P-Gal4 lines allows obtaining Gal4 insertions in the bithorax complex of Drosophila. Mech Dev 123: 860-867.

66. Ahmad SM, Baker BS (2002) Sex-specific deployment of FGF signaling in Drosophila recruits mesodermal cells into the male genital imaginal disc. Cell 109: 651-661.

67. Bienz M, Saari G, Tremml G, Muller J, Zust B, et al. (1988) Differential regulation of Ultrabithorax in two germ layers of Drosophila. Cell 53: 567-576.

68. Freeland DE, Kuhn DT (1996) Expression patterns of developmental genes reveal segment and parasegment organization of D. melanogaster genital discs. Mech Dev 56: 61-72.

69. Damen WG, Tautz D (1999) Abdominal-B expression in a spider suggests a general role for Abdominal-B in specifying the genital structure. J Exp Zool 285: 85-91.

70. Dolle P, Izpisua-Belmonte JC, Brown JM, Tickle C, Duboule D (1991) HOX-4 genes and the morphogenesis of mammalian genitalia. Genes Dev 5: 1767-1767.

71. Warot X, Fromental-Ramain C, Fraulob V, Chambon P, Dolle P (1997) Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. Development 124: 4781-4791.

72. van der Hoeven F, Sordino P, Fraudeau N, Izpisua-Belmonte JC, Duboule D (1996) Teleost HoxD and HoxA genes: comparison with tetrapods and functional evolution of the HOXD complex. Mech Dev 54: 9-21.

73. Kelsh R, Dawson I, Akam M (1993) An analysis of abdominal-B expression in the locust Schistocerca gregaria. Development 117: 293-305.

74. Huang HF, Li MT, Von Hagen S, Zhang YF, Irwin RJ (1997) Androgen modulation of the messenger ribonucleic acid of retinoic acid receptors in the prostate, seminal vesicles, and kidney in the rat. Endocrinology 138: 553-559.

75. Thomson AA, Marker PC (2006) Branching morphogenesis in the prostate gland and seminal vesicles. Differentiation 74: 382-392.

76. Hendrickson JE, Sakonju S (1995) Cis and trans interactions between the iab regulatory regions and abdominal-A and abdominal-B in Drosophila melanogaster. Genetics 139: 835-848 Issn: 0016-6731.

77. Hopmann R, Duncan D, Duncan I (1995) Transvection in the iab-5,6,7 region of the bithorax complex of Drosophila: homology independent interactions in trans. Genetics 139: 815-833 Issn: 0016-6731.

78. Minami R, Wakabayashi M, Sugimori S, Taniguchi K, Kokuryo A, et al. (2012) The homeodomain protein defective proventriculus is essential for male accessory gland development to enhance fecundity in Drosophila. PLoS One 7: e32302.

79. Gong WJ, Golic KG (2004) Genomic deletions of the Drosophila melanogaster Hsp70 genes. Genetics 168: 1467-1476.

80. Patton JS, Gomes XV, Geyer PK (1992) Position-independent germline transformation in Drosophila using a cuticle pigmentation gene as a selectable marker. Nucleic Acids Res 20: 5859-5860.

81. Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, et al. (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73: 56-65.

82. Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG (2005) Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res 33: e36.

83. Smith D, Wohlgemuth J, Calvi BR, Franklin I, Gelbart WM (1993) hobo enhancer trapping mutagenesis in Drosophila reveals an insertion specificity different from P elements. Genetics 135: 1063-1076.

84. Drosophila protocols -Sullivan W, Ashburner, M. and Scott, H. (2000) Cold Spring harbor Laboratory Press: p353.

85. Apitz H, Kambacheld M, Hohne M, Ramos RG, Straube A, et al. (2004) Identification of regulatory modules mediating specific expression of the roughest gene in Drosophila melanogaster. Dev Genes Evol 214: 453-459.

86. Venken KJ, He Y, Hoskins RA, Bellen HJ (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science 314: 1747-1751.

87. Castelli-Gair J, Greig S, Micklem G, Akam M (1994) Dissecting the temporal requirements for homeotic gene function. Development 120: 1983-1995.

88. Neubaum DM, Wolfner MF (1999) Mated Drosophila melanogaster females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153: 845-857.

89. Mueller JL, Linklater JR, Ravi Ram K, Chapman T, Wolfner MF (2008) Targeted gene deletion and phenotypic analysis of the Drosophila melanogaster seminal fluid protease inhibitor Acp62F. Genetics 178: 1605-1614.

CHAPTER 6

RNA-SEQ ANALYSIS OF *ABD-B* MUTANT, *IAB-6^{COCU}*, REVEALS INDIVIDUAL GENES IN THE SECONDARY CELLS OF THE MALE ACCESSORY GLAND THAT ARE ESSENTIAL FOR THE FEMALE POST MATING RESPONSE[6]

6.1 INTRODUCTION

In Chapter 5, I showed that the two cell types of the *Drosophila* male accessory gland together make products that maintain the LTR. More specifically, our and previous work suggests that while main cells produce proteins that both initiate and maintain the PMR [1,2] the secondary cells provide products that are needed to maintain the response [3]. Independent work on the accessory gland involving mutations in the Hox gene *defective proventriculus (dve),* which was found to be important for secondary cell formation upstream of *Abd-B*, supports the role of these cells in regulating long-term egg-laying [4]. Defects in long term receptivity suppression were also seen in mates of males with reduced expression of bone morphogenic protein (BMP) in the secondary cells [5]. While together these studies confirm that the secondary cells are important for regulating post mating effects, they do not address exactly what the secondary cells contribute to regulation of the LTR, beyond that normal growth and differentiation is required for their function.

Use of the *iab-6*-GAL4 driver to selectively knockdown LTR genes in the secondary cells revealed that they produce three LTR network proteins (CG1656, CG1652, and CG17575) [3,6]. However, none of these proteins were absent in the *iab-6^{cocu}* mutant, suggesting that

[6] This project is ongoing and done in collaboration with D. Gligorov and François Karch at the University of Geneva. While I Consulted with them on the design of the RNA-seq experiment the bulk of that experiment and the morphological work was conducted by them. Analysis of the RNA-eq results and subsequent candidate selection and experimental data presented in this chapter is the work of J.Sitnik.

additional as yet unknown contributions from this cell type may be essential for the LTR. Further, it is unclear how, if at all, the defects in cellular morphology and differences in glycosylation, stability, and abundance of known LTR proteins observed in *iab-6*^{cocu} mutants relate to the maintenance of the PMR.

To determine the impact of the *iab-6*^{cocu} deletion on gene expression and to identify individual genes that underlie the defects seen in the LTR, our collaborators at the University of Geneva used RNA-seq to compare transcripts from *iab-6*^{cocu} and wild type accessory glands. I selected 19 down-regulated genes for knockdown using secondary cell specific RNAi. Of the secondary cell genes, my tests identified 8 whose expression in the secondary cells is necessary for normal long-term egg-laying , 7 of which also are needed to regulate long-term female receptivity. Surprisingly, only one of the 8 genes encodes a protein known to be transferred to females, CG3349. One explanation for this finding is that the *iab-6*^{cocu} mutant may not primarily affect Sfps directly but instead might work through disrupting other cellular functions such as vacuole associated secretion. This work is still ongoing and I have yet to connect differences in observed protein stability or abundance to the maintenance of the LTR.

6.2 RESULTS

RNA-seq reveals 73 candidate genes down-regulated by more than 5-fold

In the hopes of expanding our understanding of secondary cell function as well as to identify genes impacted by the *iab-6*^{cocu} mutation our collaborators in the Karch Lab performed an mRNA-seq analysis comparing the transcriptomes of wild-type accessory glands and *iab-6*^{cocu} accessory glands. To control for background and the PhiC31 insertion used to make the *iab-6*^{cocu} mutant (see Chapter 5) they used the *iab-5,6*^{rescue} line, which integrates a wild type copy of *Abd-B* instead, as the control flies. The HiSeq run, conducted as a single replicate for screening

purposes (See METHODS), yielded 66,943,897 reads, 36,740,061 for *iab-5,6*^{rescue} and 30,203,836 for *iab-6*^{cocu}. For both samples approximately 71% of the reads were mapped onto the reference genome, resulting in 8,764 genes. Fold differences were calculated by dividing the number of reads per gene (*iab-5,6*^{rescue}/ *iab-6*^{cocu}) for "down-regulated genes" (those which required *Abd-B* for high expression in the accessory gland) and the inverse for up-regulated genes (those whose expression is repressed by *Abd-B*). Below I report my analysis of the mRNA-seq data obtained by the Karch Lab.

Of the 8764 genes detected by RNA-seq, 694 were differentially expressed (up or down-regulated) by at least 2 fold, suggesting that loss of *Abd-B* expression in the secondary cells has a substantial impact on the accessory gland. Surprisingly, the majority of these expression differences represented increases in expression in *iab-6*^{cocu} mutant relative to controls (Summary Table 6.1). Although one explanation could be that *iab-6*^{cocu} causes a homeotic transformation of the secondary cells to main cells, our previous data (see Chapter 5, [3]) does not support this hypothesis. At this time we cannot differentiate between several hypothesis for this increase: (1) ectopic expression of main cell transcripts in the secondary cells, (2) an increase in expression in the main cells, or (3) increased expression of transcripts already present in the secondary cells due to loss of repression by *Abd-B* in the *iab-6*^{cocu} mutant. No difference in >2-fold expression between *iab-6*^{cocu} and control accessory gland samples was detected for known LTR proteins regardless of cellular origin (Table 6.2), consistent with our conclusion that loss of known LTR proteins does not underlie the phenotypes observed in mates of *iab-6*^{cocu} males (Table 6.2). However, it is important to note that genes expressed equally in both cell types would not be detected in this screen due to the small relative number of secondary cells per accessory gland (roughly 4% of the gland). Of the 694 differentially expressed genes, only 14 were previously

Table 6.1: Summary of differentially expressed genes in *iab-6*^{cocu} males

			Up regulated			Down regulated		
	Total	**Total >2**	**>5**	**>2**	**Acp**	**>5**	**>2**	**Acp**
# Genes	8764	694	115	433	6	73	261	8

Table 6.2: Expression of genes encoding known LTR proteins in *iab-6*^{cocu} males

Known LTR Proteins					
Gene	**Fold Change**	**Direction of Change**	**Chromosome**	**Feature**	**Cellular Origin**
CG1656[1]	1.27	+	2R	Lectin	Secondary[3]
CG1652[1]	1.32	+	2R	Lectin	Secondary[3]
CG9997[1]	1.04	-	3R	Protease	Main[3]
CG17575[1]	1.72	+	2R	CRISP	Secondary[3]
Seminase[8]	1.10	-	3L	Protease	Main[3]
Sex Peptide[6,7]	1.08	-	3L	peptide/prohormone	Main[7]
Intrepid	1.05	-	3R	Protease	?
Antares	1.05	+	2R	CRISP	?
Aquarius	1.03	-	3R	Protease	?

The expression differences for known Long-term response genes [1,3,6,7,8]. Intrepid, Antares, and Aquarius are discussed in Chapter 4. None of these genes show differential expression >2-fold suggesting that they are not impacted by Abd-B expression.

identified as encoding accessory gland proteins (Table 6.3). This suggests that the consequences of the *iab-6*^{cocu} mutation may not be confined to the production of specific Sfps but may instead impact other cellular functions necessary for the PMR. This is consistent with the cellular morphology defects observed in *iab-6*^{cocu} mutants, specifically loss of the large vacuoles characteristic of the cell type which may impair protein storage or transport in these cells, as well as the observation that the *iab-6*^{cocu} mutation impacts the glycosylation state of LTR associated Sfps CG1656, CG1652, and CG17575 [3]. To both reduce the candidate pool, and the likelihood of false hits, I used a more stringent arbitrary cut-off value of >5-fold change in expression before candidate selection. Further, I focused my studies in this chapter on down regulated genes, in part because I can reduce expression of these genes easily by using the UAS-GAL4 system to drive knockdown in the secondary cells using lines from the Vienna Stock Center [9].

In total, 73 genes were down-regulated in *iab-6*^{cocu} flies by a factor of five or more relative to *iab-5,6*^{rescue} flies (Table 6.4). Based on Fly Atlas expression data [10], very few of our down regulated genes are primarily or exclusively expressed in the accessory gland (AG) (Table 6.5). Rather, 24 (32.8%) show highest expression in the Malpighian tubules, 12 (16.4%) in the testis, 9 (12.3%) in the midgut, and only 2 (2.7%) in the AG. These two AG specific genes are CG11598 and CG3349, two previously identified Sfps [11] that have not yet been characterized. However, 38 (52.02%) of the 73 genes have a signal sequence, suggesting that they could potentially encode Sfps that were not detected in previous assays either due to rarity or other factors.

I was initially concerned with the high number of genes with broad expression patterns or that are primarily expressed in the Malpighian tubules in our down-regulation candidates. The Malpighian tubules often wrap around the male reproductive tract during dissection, making

Table 6.3: Differential expression of known Sfps in *iab-6*^{cocu} males

Known Accessory Gland Proteins				
Gene	**Fold Change**	**Direction of Change**	**Chromosome**	**Feature**
CG11598	66.1	-	3R	Lipase
CG13309	26.3	-	3L	Unknown
scpr-A	14.3	-	3R	CRISP
CG3349	5.8	-	3L	Unknown
Cdlc2	5.1	-	2L	Dynein light chain
Obp51a	4.3	-	2R	Odorant Binding
Peritrophin-A	3.6	-	X	Chitin Binding
CG14913	3.0	-	2L	Unknown
Obp56g	72.6	+	2R	Odorant Binding
Acp24A4	53.9	+	2L	Protease Inhibitor
CG6426	6.2	+	2R	Destabilase
Glt	5.8	+	2L	Carboxylesterase
Cpr67B	5.7	+	3L	Cuticle Protein
Phm	2.6	+	2R	Monooxygenase

Table 6.4: Genes with down-regulated expression in *iab-6*^{cocu} males						
Gene ID	**Fold Decrease**	**Signal Sequence**	**Predicted function**	**Primary Expression**	**AG expression**	**Male biased**
alpha-Est5	25.48	No	Carboxylesterase	Midgut	none	No
alpha-Est7	21.92	No	Carboxylesterase	Broad	moderate	Yes
alphaTry	-	Yes	trypsin, serine protease	Midgut	none	Yes
beat-Ic	5.98	No	Cell Adhesion	CNS	low	Yes
beat-IIIa	5.17	No	immunoglobulin-like	Broad	none	No
beat-IV	-	Yes	immunoglobulin-like	CNS	none	Yes
betaTry	-	Yes	trypsin, serine protease	Midgut	none	No
Cdlc2	5.14	No	Dynein light chain	Testis	none	Yes
CG10514	25.48	No	CHK-kinase like	Malpighian Tubules	none	Yes
CG10560	-	No	CHK-kinase like	Malpighian Tubules	none	Yes
CG10764	6.58	Yes	Peptidase S1A	Testis	none	Yes
CG11598	66.11	Yes	Lipase	Accessory Gland	moderate	Yes
CG11892	11.51	No	CHK kinase-like, DUF227	Malpighian Tubules	none	Yes
CG12374	-	Yes	M14 Protease inhibitor	Midgut	none	No
CG12506	11.30	Yes	Unknown	Testis	none	Yes
CG13309	26.31	Yes	Unknown	Malpighian Tubules	low	Yes
CG13538	41.10	No	Unknown	Testis	none	Yes
CG13793	32.21	No	Neurotransmitter	Heart	low	Yes
CG13830	-	No	Peptidase M20	Malpighian Tubules	none	No
CG14069	-	Yes	Unknown	Testis	none	Yes
CG14245	9.04	Yes	Unknown	-	-	Yes
CG14246	9.04	Yes	Unknown	Malpighian Tubules	none	Yes
CG14292	18.09	Yes	Unknown	Malpighian Tubules	low	No
CG14376	41.93	Yes	solute-binding protein	Broad	low	No
CG14715	7.35	Yes	peptidyl-prolyl cis-trans isomerase, protein folding	Broad	moderate	female
CG15155	20.55	No	Acyl-CoA N-Acyltransferase	Malpighian Tubules	none	No
CG15406	-	Yes	general/sugar transporter	Malpighian Tubules	none	Yes
CG15614	12.33	No	G-coupled protein receptor	Broad	none	No
CG17752	-	Yes	transmembrane transport	Malpighian Tubules	none	Yes
CG18088	5.49	Yes	Alkaline Phosphatase	salivary gland	low	female
CG2187	18.09	No	sodium symporter	Malpighian Tubules	low	No
CG2196	-	No	transmembrane transport	Malpighian Tubules	low	No
CG3106	10.41	Yes	Acyl-CoA N-Acyltransferase	Midgut	none	No
CG31090	29.60	No	sodium symporter	Malpighian Tubules	none	Yes
CG31198	17.26	Yes	Peptidase M1	Midgut	none	No
CG31272	-	No	Lipid Transport	Malpighian Tubules	low	No
CG31388	20.00	Yes	CAP domain protein, allergen	Testis	none	Yes
CG3285	-	Yes	general/sugar transporter	Malpighian Tubules	none	Yes
CG3349	5.77	Yes	Unknown	Accessory Gland	moderate	Yes
CG33630	16.44	Yes	Unknown	Broad	low	No

Table 6.4: Genes with down-regulated expression in *iab-6*^{cocu} males						
Gene ID	**Fold Decrease**	**Signal Sequence**	**Predicted function**	**Primary Expression**	**AG expression**	**Male biased**
CG33631	160.31	Yes	Unknown	Broad	none	-
CG33775	-	No	Unknown	Broad	none	No
CG33783	1043.24	No	Unknown	-	-	No
CG33784	573.00	Yes	Unknown	Eye	none	Yes
CG34167	21.20	No	Unknown	Testis	low	Yes
CG34366	5.75	No	voltage dependant potassium channel	Broad	-	Yes
CG3690	5.12	No	general/sugar transporter	Malpighian Tubules	low	Yes
CG41443	19.73	No	Unknown	-	-	-
CG5361	22.20	No	Alkaline Phosphatase	Malpighian Tubules	low	Yes
CG6602	8.49	Yes	Unknown	Malpighian Tubules	low	Yes
CG7874	39.16	Yes	chitin binding peritrophin-A, mucin	Malpighian Tubules	low	No
CG7882	30.01	No	transporter, general sugar	Malpighian Tubules	none	Yes
CG8157	7.03	Yes	Unknown	Broad	none	No
CG8197	-	No	Unknown	Testis	none	Yes
CG9259	53.44	No	CHK-kinase like/Duff227	Malpighian Tubules	low	Yes
CG9294	34.53	No	peptidase S1/S6	Broad	none	No
CG9509	26.31	No	glucose-methanol-choline oxioreductase	Malpighian Tubules	none	Yes
Cpr56F	52.61	Yes	insect cuticle protein	Testis	none	Yes
Cyp6a14	35.68	Yes	cytochrome P450	Crop	low	Yes
dpr10	20.55	Yes	CAP domain protein	Testis	none	Yes
Jon65Aiv	21.37	Yes	peptidase S1/S6	midgut	none	No
Jon74E	-	No	Peptidase S1A	Midgut	low	No
Jon99Cii	-	Yes	Peptidase S1A	-	-	No
NaPi-T	8.02	No	Phosphate transport	Malpighian Tubules	none	Yes
nerfin-2	-	No	zinc/nucleic acid binding	CNS	none	Yes
obst-A	23.02	Yes	chitin-binding	Broad	low	No
phr	11.21	No	DNA photolase, DNA repair	Broad	low	Female
ple	6.85	No	Tyrosine-3 monohydroxilase	CNS	none	No
scpr-A	14.28	Yes	CAP domain protein	Testis	none	Yes
scpr-C	25.81	Yes	unknown, CAP allergen	Testis	none	Yes
Skeletor	28.02	Yes	Spindle Assembly	-	-	-
Traf1	14.83	No	JNK, Cell Death, Apoptosis, Cell Fate Determination	CNS	none	No
Ugt86Dj	46.04	Yes	UDP-glucosyltransferase	Midgut	low	No
Uro	9.45	No	Uricase	Malpighian Tubules	low	No

Table compiled from signal sequence predictions obtained from SignalP 4.0 [12], annotated functions from FlyBase [13], primary and accessory gland expression from Fly Atlas [10], and male biased data from modENCODE [14,15]. Biased expression was set at a 2-fold difference in expression based on whole fly data from modENCODE (ages 1d, 5d, and 30d). No fold differences were able to be calculated for entries with (-) since the gene was not detected in *iab-6*^{cocu} males. Expression data delineated with a (-) was not available.

Table 6.5: Site of highest expression for down-regulated genes

Organ	# genes	%
Malpighian tubule	24	32.4%
Broad expression	13	17.6%
Testis	12	16.2%
Midgut	9	12.2%
Unknown	6	8.1%
Brain	5	6.8%
Accessory Gland	2	2.7%
Eye	1	1.4%
Salivary Gland	1	1.4%
Heart	1	1.4%
Crop	1	1.4%

Based on Fly Atlas data [10]. Broad expression indicates that there was no clear winner for highest expression.

them a likely source of sample contamination. Further, I am looking for changes in expression based on a small number of cells, roughly 4% of the accessory gland. A low level of contamination in the control sample and absence of contamination in the *iab-6*^{cocu} sample would mimic the down-regulation I am trying to detect. Since I have access to only one replicate of mRNA-seq data I attempted to rule out possible contaminants using other criteria. One of the down-regulated genes, the known Sfp CG13309 [16], is most highly expressed in the malpighian tubules. However, CG13309 also has male-biased expression [14,15] and is expressed at very low levels in the accessory gland [10]. Of the possible contaminants, 25/46 of these genes also show male biased expression [15] suggesting that they may have sex specific functions and 17/46 (5/25 that also have male biased expression) have previously been found to be expressed in the accessory gland . As such, despite my initial concerns, I kept these 37 genes in my original candidate pool and removed the 9 genes that shared none of these features with CG13309.

From the remaining pool of 64 down-regulated genes I prioritized candidates (in order) based on availability of fly lines, predicted function, the presence of a signal sequence, and confirmed (though not necessarily exclusive) expression in the AG. Because some of the phenotypes observed in *iab-6*^{cocu} males appear to be post-translational (such as protein processing and glycosylation [3]), I preferentially chose genes with predicted functions related to modification of proteins (such as proteases or phosphatses) or that played a role in sugar transport, binding, or transfer. In addition, since the vacuole is absent in *iab-6*^{cocu} males, I included other transport proteins that may be important for vacuole function. For candidates that did not fall into these categories, I required the presence of a signal sequence, indicative of potential Sfps. These genes also had to have known accessory gland expression and/or show

male biased expression to be considered. In total, I selected 20 candidates (including 3 known ACPs) for testing (Table 6.6).

Receptivity suppression is compromised in mates of males knocked down for candidate genes.

Unlike mates of control males, mates of *iab-6*^{cocu} males return to virgin levels of receptivity by 4 days after the start of mating (ASM). As an initial screen for candidate genes associated with the LTR, I generated knockdown males by crossing the *iab-6D1*-GAL4 driver to flies containing a single hairpin RNA directed at my gene of interest [9]. The *iab-6D1*-GAL4 driver has an expression pattern similar to that of the *iab-6*D5-GAL4 driver used in Chapter 5 [3] (See Appendix C for expression of *iab-6D1*-GAL4), however the size of the putative enhancer sequence is 1.2kb rather than the full 2.8kb. For controls I crossed the driver to either the *Attp*^{60100} or *w*^{1118} background stock where appropriate. Knockdown for each gene is currently being assessed by RT PCR, however CG15092 was found to not knock down and will not be discussed further. Virgin Canton S females were mated to either RNAi or control males and then housed individually for 4 days after which they were allowed access to a wild type male for 1 hour. Of the 19 candidate genes tested, mates of 7 different RNAi males (CG14292, CG3349, CG15406, CG43161, CG3285, CG7882, and CG14069) were significantly more receptive than mates of control males (Figure 6.1). One of the RNAi lines that impacted 4d receptivity, CG15406, has a confirmed off-target, the gene CG31326. Knockdown of CG31326 in the secondary cells did not affect the receptivity of mated females, suggesting that differences in remating observed for the CG15406 line are either due to CG15406 knockdown or a combination of knockdown of the two genes. Only CG3349 was previously known to be a transferred Sfp [17],.

Table 6.6: RNA-seq Candidate Genes

CG#	Chromosome	ID	Fold Decrease	Signal Sequence	Function	Highest Expression	Male biased
CG14069	2L	CG14069	-	yes	unknown	Testis	Yes
CG15406	2L	CG15406	-	no	general/sugar transporter	Malpighian tubules	Yes
CG3285	2L	CG3285	-	no	general/sugar transporter	Malpighian tubules	Yes
CG33783	3R	CG33783	1043.24	no	unknown	Unknown	No
CG33784	3R	CG33784	573.00	yes	unknown	Eye	Yes
CG33631	3R	CG33631	160.31	yes	unknown	Broad	No
CG11598	3R	CG11598	66.11	yes	lipase	Accessory gland	Yes
CG9036	2R	Cpr56F	52.61	yes	cuticle protein	Testis	Yes
CG15902	3R	Ugt86Dj	46.04	yes	UDP-glucosyl transferase	Midgut	No
CG14376	3R	CG14376	41.93	yes	solute-binding protein	Broad	No
CG7882	2R	CG7882	30.01	no	transporter, general sugar	Malpighian tubules	Yes
CG43161	3R	Skeletor	28.02	yes	spindle assembly	-	-
CG13309	3L	CG13309	26.31	yes	unknown	Malpighian tubules	Yes
CG5361	3R	CG5361	22.20	no	alkaline phosphatase	Malpighian tubules	Yes
CG1112	3R	alpha-Est7	21.92	no	carboxyl-esterase	Broad	Yes
CG14292	3R	CG14292	18.09	yes	unknown	Malpighian tubules	Yes
CG33630	3R	CG33630	16.44	yes	unknown	broad	No
CG14715	3R	CG14715	7.35	yes	cis-transisomerase	Broad	Female
CG3349	3L	CG3349	5.77	yes	unknown	Accessory gland	Yes
CG18088	2L	CG18088	5.49	yes	alkaline phosphatase	Salivary gland	Female
CG32593	X	Flo-2	1.08	no	oxioreductase	Malpighian tubules	No
CG31326	3R	CG31326	0.38	yes	serine protease	Spermathecae	No

I selected candidate genes based on a conglomeration of criteria including fold change, signal sequence, predicted function, expression pattern, and sex biased expression. Particular preference was given to candidate genes that might be involved with glycosylation, sugar transport, or other post-translational modification in order to try and determine if these phenotypes in the *iab-6*^{cocu} mutant relate directly to the LTR. Fold decrease could not be calculated for CG14069, CG15406, and CG3285 since these transcripts were not detectable in *iab-6*^{cocu} males. Expression data was obtained from Fly Atlas [10] or, in the case of male biased data, from modENCODE [14,15]. No expression data was available for CG43161. Shaded genes were used as controls. CG32593 is a negative control gene for the malpighian tubule, whereas CG31326 is an off-target for the CG15406 RNAi line.

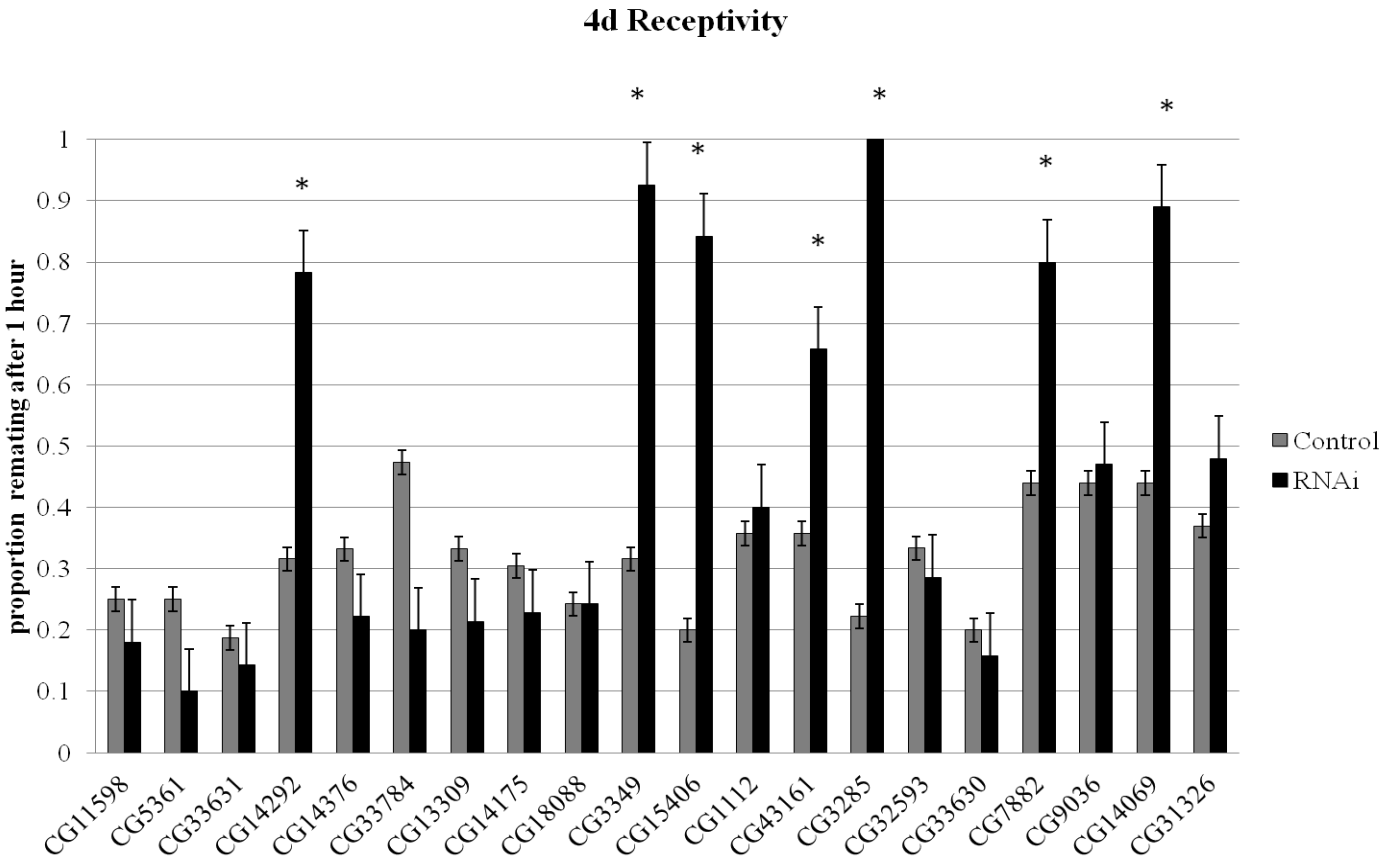


Figure 6.1: 4d receptivity in mates of RNAi candidate males

The percentage of females previously mated to either RNAi (black) or control (grey) males that were willing to remate 4ds later during exposure to a wild type male for one hour. Of the genes tested, 7 failed to sustain post mating suppression of receptivity in their mates (* denotes WRST p=<0.02, Ns ranged from 15-30 per treatment).

Three of the 7 genes (CG7882, CG3285, and CG15406) encode proteins that lack signal sequences and that are classified as sugar transporters. These proteins are most likely not secreted Sfps. This points to the possibility that the glycosylation differences observed in *iab-6*^{cocu} males may be the result of a decrease in available sugars rather than a specific problem in the glycosylation process.

Perhaps more interesting, albeit confusing, is the results for CG43161. Also known as *skeletor*, CG43161 encodes a putative component of the nuclear spindle, referred to as the spindle matrix. This matrix is thought to form before the mitotic spindle during mitosis [18] and also contains the proteins Chromator [19,20,21,22] and Megator [23,24]. The exact composition of the spindle matrix is unclear, however it is composed of filaments that do not stain with antibodies to Actin or Tubulin and that are not depolymerized when exposed to the drug nocodozole [18,25,26] suggesting that the spindle matrix does not require polymerization to form. Since there is no apparent difference in accessory gland size or secondary cell number in *iab-6*^{cocu} males it seems unlikely that cellular division is negatively impacted by loss of *Abd-B* expression in the secondary cells (Chapter 5, [3]). Electron microscopy of the secondary cells, first conducted by Bairati [25,27], noted that the vacuoles of the secondary cells and the lumen of the accessory gland contain filaments that are not composed of Actin or Tubulin [28]. The composition, source, and function of these filaments is unclear but it is possible that they could be extensions of the spindle matrix. Unlike *chromator* and *megator*, s*keletor* encodes for a protein with a predicted signal sequence [12] suggesting that it may be secreted into the lumen of the accessory gland and that it could be a potentially transferred Sfp.

The three remaining candidate genes (CG14292, CG14069, and CG3349) are of unknown function. As previously mentioned, CG3349 is known to encode a transferred Sfp.

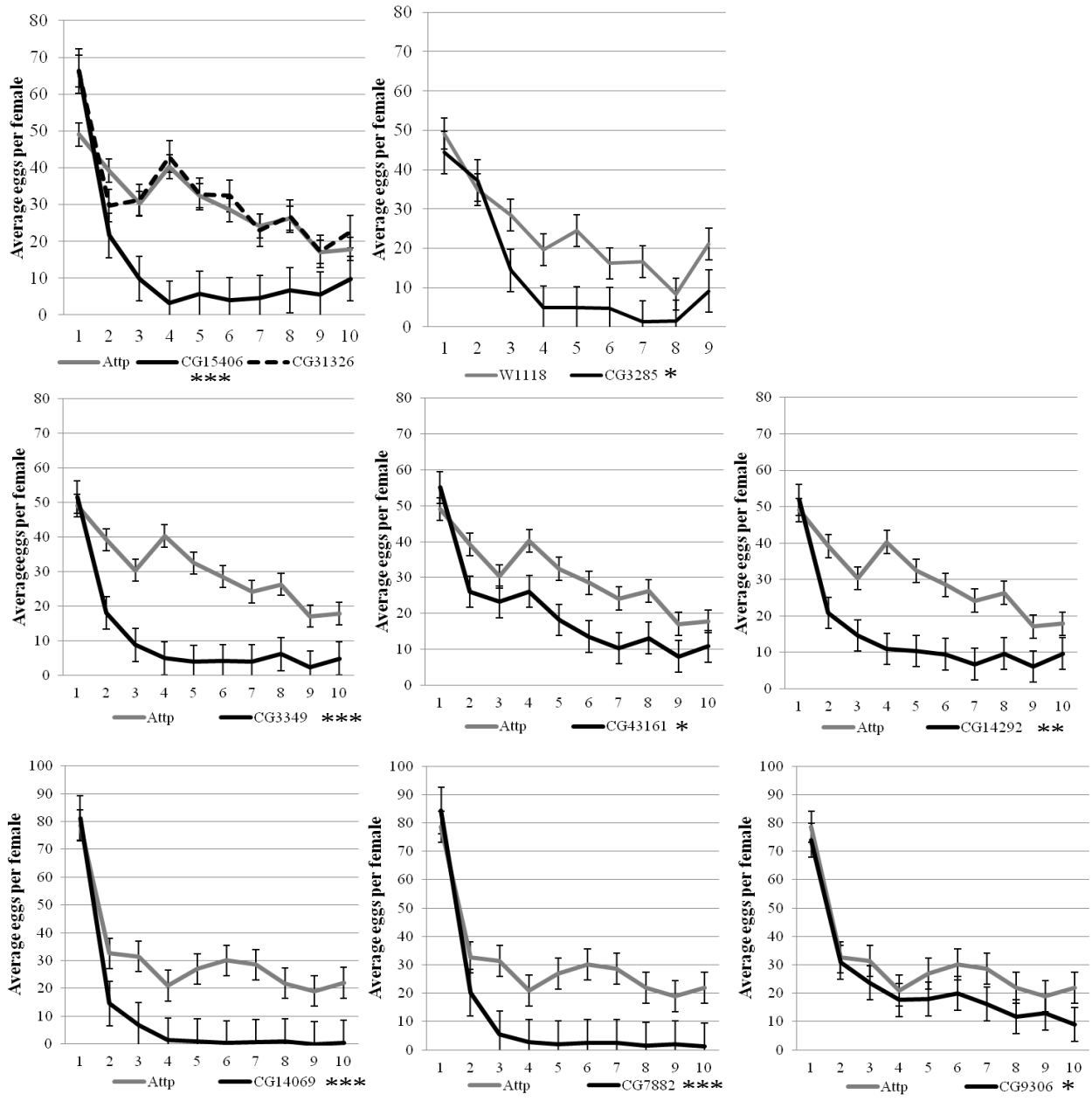
Both, CG14292 and CG14069 encode proteins that contain a signal sequence, suggesting that they may be secreted Sfps. However, none of these genes encode proteins with conserved domains [29] and thus offer no direct insights into what they might contribute to the LTR.

Mates of candidate males lay fewer eggs than mates of control males

To test if the genes uncovered in our receptivity screen also impact the ability of males to induce egg-laying in their mates, I crossed RNAi or control males to virgin females and counted the number of eggs laid over a 9-10 day period. Because my collaborators in the Karch lab independently detected some egg-laying effects for CG9036. I also included it in my assays to verify this effect, despite there being no effect on receptivity.

All 8 candidate genes play a role in maintaining long term female egg-laying (Figure 6.2). Reduced expression of CG15406 in males also increases short term egg-laying in their mates, however this increase appears to be an off-target effect of CG31326 knockdown. All of the other candidate genes elicited normal 24h egg-laying responses, consistent with my observations in the *iab-6*^{cocu} mutant and their potential role in regulating the LTR. Similarly, no significant differences in hatchability (# progeny/#eggs) were detected for any of the RNAi candidates (data not shown).

The cuticle protein CG9036 appears to only impact egg-laying and not receptivity, similar to CG32833 (Chapter 3). This is unusual, since most LTR associated proteins influence SP storage which in turn results in defects in the maintenance of both long-term egg-laying and receptivity suppression. However, these results suggest that receptivity and egg-laying may be separable. The reverse phenotype (inability to maintain receptivity reduction but normal maintenance of long-term egg-laying) has also been observed before (see Appendix C). It is



6.2: Long term egg-laying is reduced in mates of secondary cell RNAi males

The mean number of eggs laid per female mated to either control males (grey) or RNAi males (black) over a 9-10 day period. Graphs of experiments run at the same time using the same Attp control have identical control plots (Set1: CG15406, CG3349, CG43161, and CG14069 - Set2: CG14069, CG7882, and CG9306) and were regraphed for easier analysis. Mates of all 8 RNAi males lay fewer eggs over the long term than mates of control males (* p = <0.05, ** p =<0.005, *** p=<0.0005, Ns 15-26). Mates of both CG15406 and CG31326 KD males laid more eggs in the first 24 hours than mates of controls, suggesting that the off-target gene CG31326 could play a role regulating proteolysis events involved in this process.

unclear how a Sfp might function downstream of SP, since ectopic expression of SP in virgin females is sufficient to induce both responses [30]. Still, looking into the possibility of separating egg-laying increases from receptivity suppression is tantalizing and unfortunately outside the scope of this chapter.

Sex peptide storage is abnormal in mates of candidate males

The defects in long-term egg-laying and receptivity suppression seen in mates of males knocked down for CG14292, CG3349, CG3285, CG14069, CG7882, CG15406, and CG42161 suggest that these secondary cell genes may impact the LTR through effecting sex peptide storage. To test this hypothesis, this I performed Western blots using SP antibodies to detect differences in stored SP at various time points ASM. In all experimental and control matings SP was produced by the male and transferred to females. So far, mates of all of the candidate genes tested (CG14292, CG43161, CG3285, CG15406, and CG3349) show loss of SP by 4d ASM (Figure 6.3). Knockdown of the off-target gene for CG15406, CG31326, does not cause this effect suggesting that CG15406 or a combination of both genes may be necessary for SP to remain in storage. However, CG14069, CG7882, and CG9036 still need to be tested. While my results for the other 5 genes are reproducible they are not yet in a consistent format so they also bear repeating. Still, these results suggest that defects in the LTR observed in mates of these RNAi males arises from a failure to maintain SP in storage in mated females. Protein products of those genes that contain signal sequences may function in the LTR network within the mated female. This is particularly likely for CG3349, as it is known to be transferred during mating. Conversely, and perhaps probably the case for the non secreted candidates, these genes may indirectly influence the LTR by producing products necessary for proper posttranslational

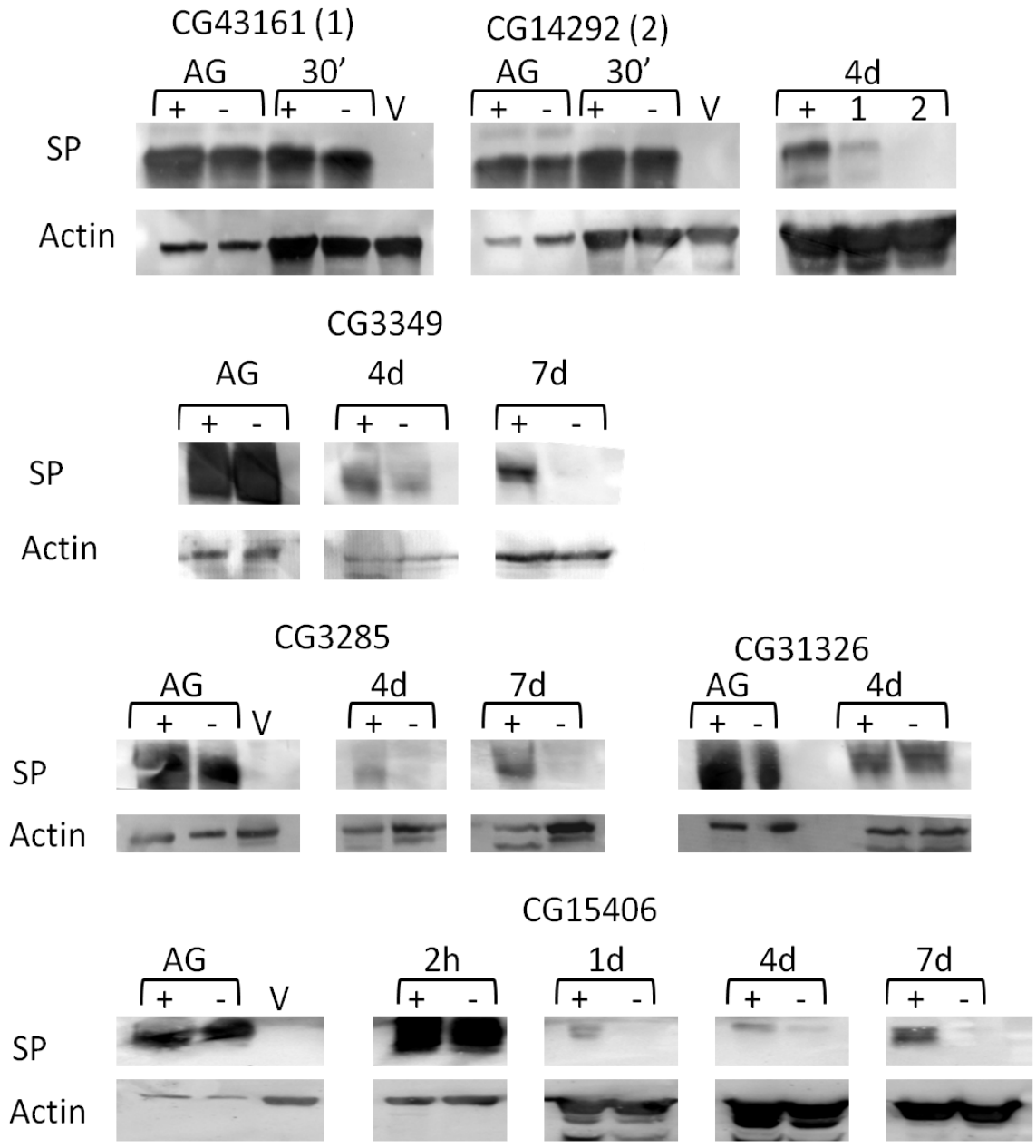


Figure 6.3: SP storage in mates of candidate RNAi males

Western blots of male accessory gland extracts (AG) and lower reproductive extracts (RTs) from mated females probed with antibodies to SP and Actin. Virgin Canton-S females were mated to either control (+) or RNAi (-) males for each gene. Male AG lanes contain ~1.5 accessory glands per lane and virgin female (V) lanes contain 4 RTs. Mated females were flash frozen at 30' (2 RTs), 2h (2 RTs), 1d (15 RTs), 4d (18 RTs), or 7d (20 RTs) ASM. All tested genes are necessary for SP storage with the exception of the off-target CG31326, a control for the candidate gene CG15406. In the blot for CG43161 and CG14292 the 4d sample is labeled with (1) for CG43161 and (2) for CG14292.

modification or secretion of LTR network proteins.

Correlation between SP storage and the stability, glycosylation, and abundance of LTR network proteins

In addition to defects in SP storage, the *iab-6*^{cocu} mutation impacts the abundance, stability, and glycosylation of LTR network proteins. The biological relevance of these changes is currently unclear. To identify individual genes responsible for these phenotypes and look for correlations between them and SP storage, I am currently testing all 19 candidate genes by Western blot using antibodies to CG1656, CG1652, CG9997, and CG17575. So far, none of the tested genes has had obvious impacts on the stability of CG9997 after transfer (Figure 6.4 &6.5C & not shown). Mates of some candidates (particularly CG11598, CG1112, CG18088, and CG33630) showed variable abundance of CG17575 at 30min ASM, with only CG33630 resulting in an increase in CG17575 abundance similar to that previously seen in mates of *iab-6*^{cocu} males (Figure 6.5 & not shown). None of these genes showed defects in receptivity suppression (Figure 6.1), suggesting that differences in CG17575 abundance may not be directly tied to the LTR phenotypes observed in mates of *iab-6*^{cocu} males. However, the effect of these genes on CG17575 abundance was variable across repeated trials, so it is difficult to definitively rule this out.

Reduced expression of two genes, CG14292 and CG14069, did consistently impact the apparent molecular weight of CG1656 and CG1652 (Figure 6.5A&B). This is surprising, because neither of these genes are predicted to play a direct role in glycosylation, sugar binding, or sugar transport. This difference in apparent molecular weight for CG1652 in knockdown versus control males was abolished after treatment with PNGase-F in the case of CG14292,

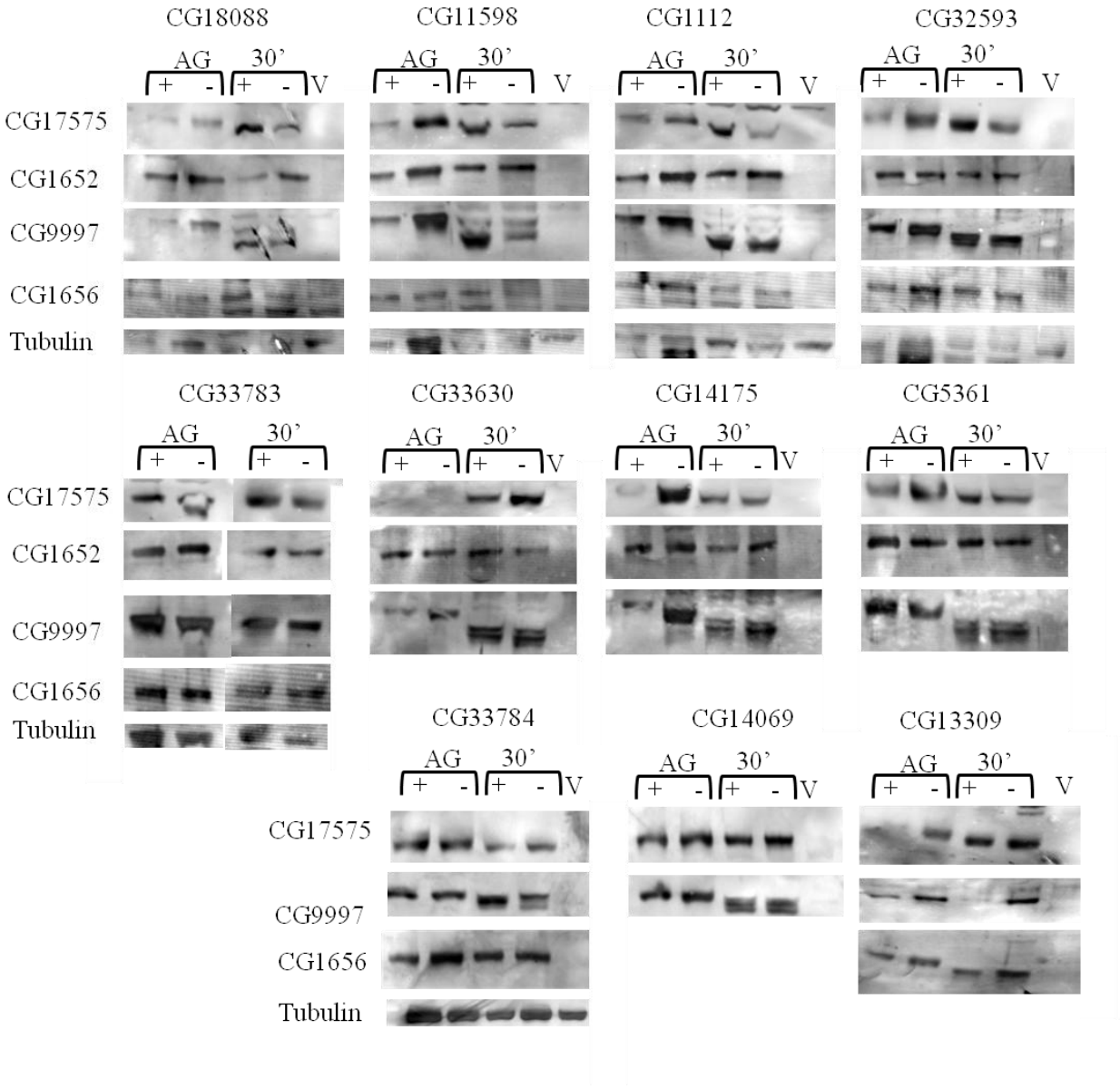


Figure 6.4: Other LTR network genes in mates of candidate males

Western blots of male accessory gland extracts (AG) and lower reproductive extracts (RTs) from mated females using antibodies to CG17575, CG1656, CG1652, CG9997, and Tubulin. Virgin females were mated to either control (+) or RNAi (-) males for each gene. Male AG lanes contain ~1.5 accessory glands per lane and virgin female (V) lanes contain 4 RTs. Mated females were flash frozen at 30' (2 RTs). No consistent differences were seen in CG9997 stability or CG1656/CG1652 glycosylation for the genes included in this figure with the exception of CG14069 (see Figure 6.5). Differences in CG17575 abundance were observed in CG18088 (decrease), CG11598 (decrease), CG1112 (decrease), CG32593 (decrease), and CG33630 (increase). These results are highly variable and need to be repeated.

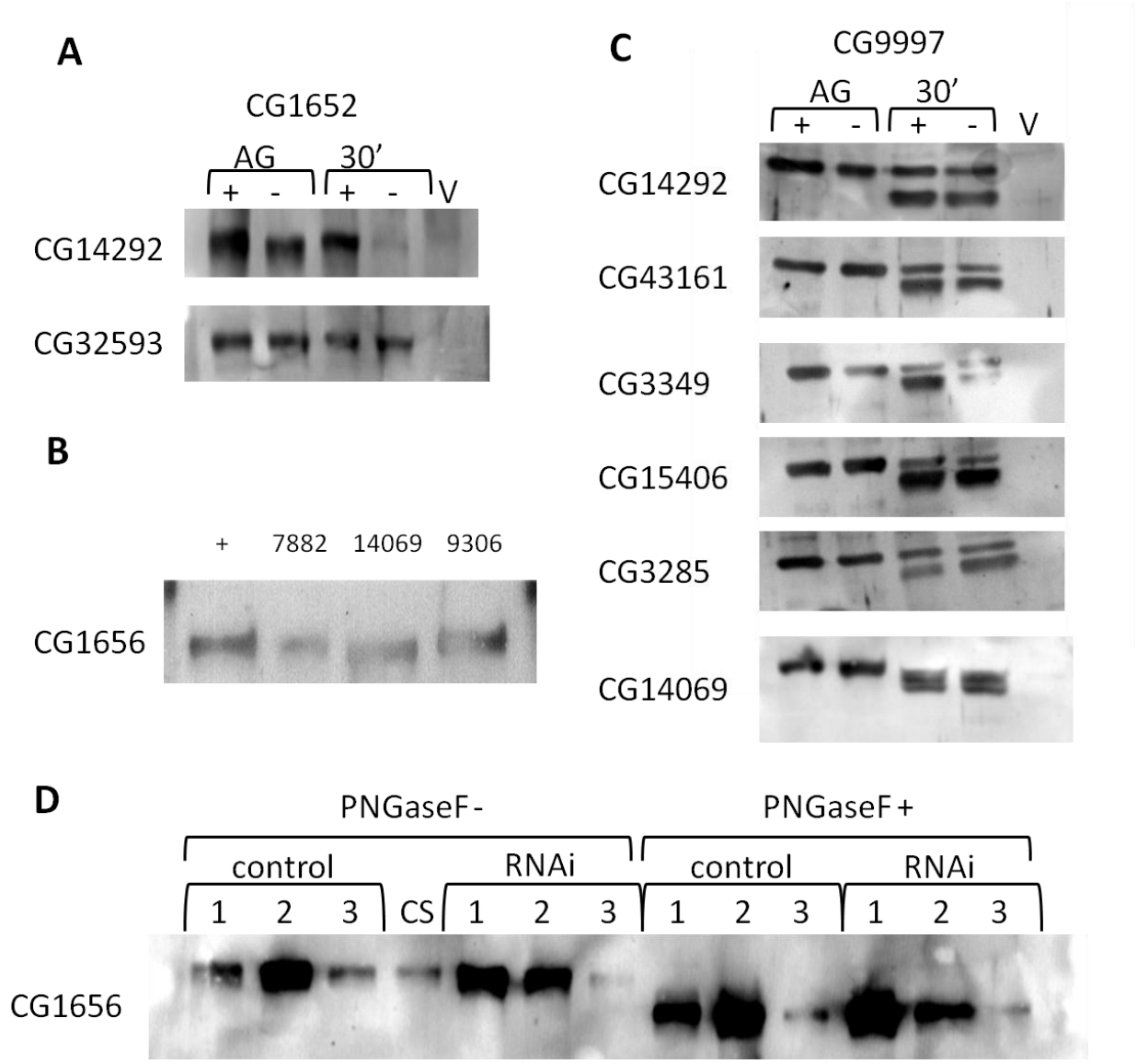


Figure 6.5: Other LTR network genes in mates of candidate males (cont.)

Western blots of male accessory gland extracts (AG) and lower reproductive extracts (RTs) from mated females using antibodies to CG9997, CG1652, CG1656, and Ovulin. Virgin females were mated to either control (+) or RNAi (-) males for each gene. Male AG lanes contain ~1.5 accessory glands per lane and virgin female (V) lanes contain 4 RTs. **A)** Knockdown of CG14292 results in a reduction in the apparent molecular weight (MW) of CG1652. The control gene CG32593 does not impact the apparent MW of CG1652. **B)** Knockdown of CG14069 reduces the apparent MW of CG1656. **C)** There is no difference in the stability of CG9997 **D)** PNGaseF treatment of accessory gland extracts (4 per lane). The difference in apparent molecular weight of CG1656 in CG14292 RNAi males is abolished with PNGaseF treatment, consistent with a difference in N-linked glycosylation.

confirming that the gel mobility disparity reflects differences in N-linked glycosylation (Figure 6.5D). CG14069 remains to be tested. Both CG14292 and CG14069 were found to be necessary for long-term receptivity suppression and egg-laying in females (Figure 6.1&2). At least in the case of CG14292, these effects appear to be caused by a reduction in stored SP in mates of CG14292 knockdown males. These results suggest that there may be a correlation between differences in the glycosylation state of CG1656/CG1652 and the storage of SP. Both CG1656 and CG1652 are necessary for the full length version of CG9997 to persist inside the female reproductive tract after mating. We originally hypothesized that the glycosylation differences in CG1656 and CG1652 detected in *iab-6*^{cocu} males may result in the loss of full length CG9997 observed in their mates. However, neither of these genes effected CG9997 processing (Figure 6.4&5) suggesting that these phenotypes may not be linked.

6.3 DISCUSSION

Using RNA-seq as a tool to screen for genes underlying the LTR defects seen in *iab-6*^{cocu} males, my collaborators identified 73candidate transcripts that appear to be down regulated by at least 5-fold in the absence of *Abd-B* expression. I examined these genes and selected 20 candidates based on expression patterns, predicted function, or the presence of a signal sequence and found that at least 8 genes (CG14292, CG3349, CG14069, CG7882, CG9036, CG15406, and CG3285) are required for the LTR. While all 8 impact egg-laying, only 7 affect receptivity suggesting that these two phenotypes may be separable. In all cases tested so far (5 of 8), reduced expression of any single one of these genes was enough to result in a reduction in stored SP.

In the original *iab-6*^{cocu} mutant, we observed abnormalities in the processing, glycosylation state, or abundance of 4 known LTR network genes (CG1656, CG1652, CG9997, and CG17575) [3]. I examined the impact of knocking down each of the 19 candidate genes on these LTR network proteins in an attempt to draw connections between these *iab-6*^{cocu} phenotypes and maintenance of the LTR None of the RNA-seq candidates were individually necessary for the processing of CG9997 or consistently impacted the abundance of CG17575 in mated females. As such, the function of CG9997 processing in the PMR, if any, is still unclear, as is the importance of these CG17575 abundance differences. However, two of the genes tested, CG14292 and CG14069, impact the glycosylation of CG1656 and CG1652. In addition, both CG14292 and CG14069 are necessary for the maintenance of the LTR. It is unclear whether or not these differences in glycosylation are causal or impact the functionality of either protein, however these results suggest that secondary cell controlled post-translational modification of these proteins via N-linked glycosylation and maintenance of the LTR are related.

Multiple process are altered in the iab-6^{cocu} mutant

Based on our results, it is clear that the *iab-6*^{cocu} mutation impacts a wide array of systems in the secondary cell that have consequences for the LTR. The *iab-6*^{cocu} mutation negatively impacts the expression of at least one transferred Sfp that is necessary for the LTR, CG3349 [31]. CG3349 has no conserved domains and its function and place in the LTR network is unclear. Four other genes necessary for the LTR, that are down-regulated in the *iab-6*^{cocu} mutant, encode secreted proteins, (CG14292, CG9036, CG43161, and CG14069). These proteins are potentially transferred Sfps but are not currently known to be transferred to females [31]. This could be due to their generally low expression in this organ and/or a lack of sensitivity in

previous assays. Knockdown of two of these genes (CG14292 and CG14069) also impacted the glycosylation of CG1656 and CG1652. Neither CG14292 nor CG14069 are predicted to be directly involved in glycosylation or sugar transport, and it is unclear whether their importance for the normal glycosylation of CG1656/CG1652 is directly linked to their impact on the LTR.

CG43161 (*skeletor*), encodes a secreted protein that is a component of the mitotic spindle matrix. Electron microscopy has shown that unidentified filaments are detectable inside the vacuoles of the secondary cell as well as in the lumen of the accessory gland [25,27]. Antigenic labeling and immunoelectrophoresis indicate that these filaments do not contain tubulin [28] but in some species of Drosophila have been shown to associate with RNA [32]. This is consistent with the observation that the spindle matrix is not composed of Actin or Tubulin [26]. Further, the filaments are detectable in mated but not virgin females and enter the seminal receptacle [27] suggesting that they are transferred during mating. Sfps, such as SP are capable of binding to sperm tails [33] leaving the possibility that these filaments may also be bound by Sfps. My results suggest that CG43161 is important in the secondary cells for the long term maintenance of the LTR. Since CG43161 is a secreted protein, it could be a transferred Sfp, similar to other LTR network proteins, independent of these filamentous structures. Conversely, CG43161 may localize to or be a component of the filaments, If so, my results would suggest a role for these filaments in regulating female post mating responses. Further work is needed to tease apart these two possibilities.

The remaining genes, CG15406, CG7882, and CG3285, do not encode for secreted proteins. Instead, these genes are predicted to encode sugar transport proteins. Sugar transport proteins are active in male reproductive tissues in mammals [34], where sugars like fructose are known to be essential in the ejaculate for sperm motility [35]. Further, fructose is secreted by the

seminal vesicles in mammals and levels of fructose in the ejaculate are used clinically as an indicator of male fertility [35,36,37,38]. Currently it is unclear whether fructose or other sugars are present in *Drosophila* ejaculates, however they are present and essential in the honey bee *Apis .mellifera* [39,40,41]. The importance of these sugar transport proteins in the secondary cells is currently unclear. They could be essential for the metabolism of the cell, and thus indirectly effecting the LTR. Conversely, the secondary cell may prove to be a source of seminal sugars similar to that of the seminal vesicle in other organisms. Still, my results indicate that knocking down any one of these three sugar transport genes in the secondary cells interferes with the maintenance of the LTR,

In addition to my analysis, my collaborators D. Gligorov and Francois Karch have identified several genes down-regulated in *iab-6*^{cocu} males that when knocked down impact cellular morphology. Two of these genes, CG14069 and CG7882, were included in my analysis and knockdown of these genes impairs the maintenance of the LTR. So far, none of the other genes I have tested were found to impact cellular morphology. The Karch lab also identified at least two other genes whose expression is necessary for normal cellular morphology but which do not appear to impact long-term egg-laying (personal correspondence with D. Gligorov, data not shown).

Conclusion

Previously we observed that the *iab-6*^{cocu} mutation influences post-translational modification of some Sfps, abnormal cellular morphology, and the maintenance of the LTR. I have demonstrated that at least 8 individual genes down-regulated in *iab-6*^{cocu} males underlie the defects observed in the LTR. Of the 8 genes identified to be essential for the LTR, 5 affect the

LTR only (CG3349, CG3285, CG9036, CG15406, and CG43161), 1 also affects glycosylation (CG14292), one affects cellular morphology (CG7882), and one affects both cellular morphology and glycosylation (CG14069) (Figure 6.7). These results suggest that these phenotypes are interrelated (as there is overlap) but genetically separable. I have yet to determine whether CG9997 processing or the increase in the abundance of CG17575 observed in *iab-6*^{cocu} males relates correlates with defects in the LTR. This work is ongoing, the current status of the project is in Table 6.5.

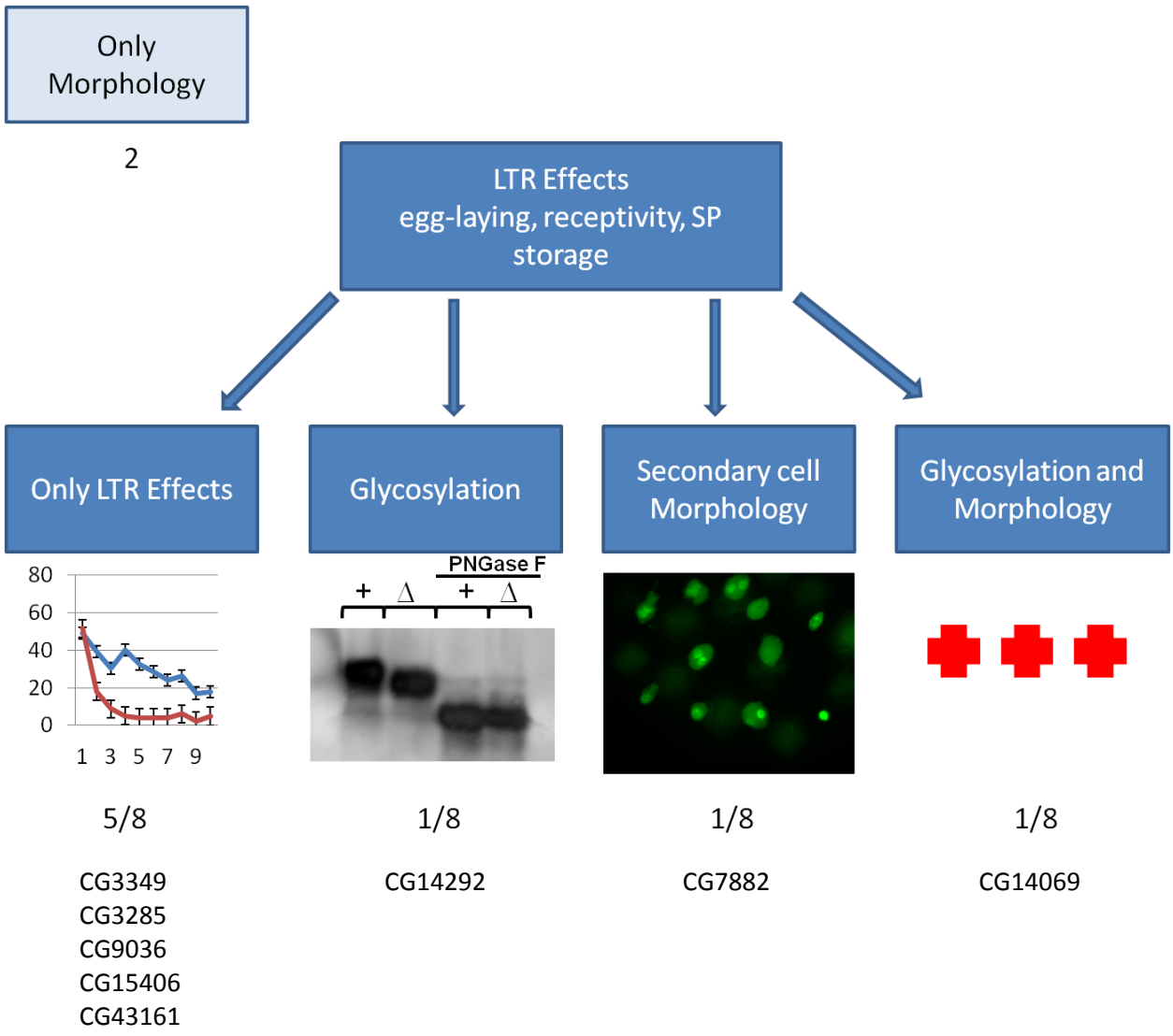


Figure 6.7: Summary of the LTR genes identified in *iab-6*^{cocu} males

A summary of the findings in this chapter. Of the 8 genes that were down-regulated in *iab-6*^{cocu} males and that were identified to be essential for the LTR, 5 affect the LTR only (CG3349, CG3285, CG9036, CG15406, and CG43161), 1 also affects glycosylation (CG14292), one affects cellular morphology (CG7882), and one affects both cellular morphology and glycosylation (CG14069). In addition, my collaborators identified 2 genes that have cellular morphology defects but do not impact the LTR (among other genes not included in this chapter). These results suggest that each of these phenotypes are interrelated (as there is overlap) but genetically separable.

Table 6.7: Current progress on *iab-6*^{cocu} RNA-seq candidate genes

Gene	4d Receptivity	Egg-laying	SP storage	CG1656/CG1652 glycosylation	17575 abundance	9997 stability	PCR
CG31326	-	-	-			-	KD
CG14292	+	+	+	+		-	
CG3349	+	+	+	-		-	
CG15406	+	+		-		-	KD
CG43161	+	+	+	-		-	KD
CG3285	+	+	+	-		-	
CG14069	+	+		+	-		
CG7882	+	+		-			
CG9036	-	+		-			
CG11598	-	n/a	n/a	-	decrease	-	
CG5361	-	n/a	n/a	-	-	-	
CG15902	-	n/a	n/a	-	-	-	Not KD
CG33631	-	n/a	n/a			-	
CG14376	-	n/a	n/a	-			
CG33784	-	n/a	n/a			-	
CG13309	-	n/a	n/a	-	-	-	
CG14715	-	n/a	n/a	-	-	-	KD
CG18088	-	n/a	n/a	-	decrease	-	
CG1112	-	n/a	n/a		decrease		KD
CG32593	-	n/a	n/a	-	decrease	-	
CG33630	-	n/a	n/a	-	+	-	KD
CG33783	-	n/a	n/a	-	-	-	

This table demonstrates the current status of the RNA-seq project outlined in Chapter 6. The assays are listed across the top: + indicates that knockdown of that gene resulted in a significant LTR phenotype similar to what was reported in Chapter 5 for *iab-6*^{cocu}, - indicates no difference between knockdowns or controls. Red cells indicate that there are insufficient data for this assay either because the assay has not yet been performed for that gene or because the result is not interpretable. The column labeled FFA stands for Fertility/Fecundity Assay and refers to egg-laying. Most of the remaining work is RT-PCR for testing knockdown, which has proven difficult due to the low expression of some of these genes in the accessory gland and the high expression they have in other tissues. CG17575 abundance has also proven difficult, due to high variability.

6.4 MATERIALS AND METHODS

mRNA-seq

The mRNA seq experiment was conducted by our collaborators, D. Gligorov and F. Karch at the University of Geneva. Total RNA was isolated from 100 pairs of accessory gland per genotype from *iab5,6*^{rescue} and *iab-6*$^{\Delta 5}$males using the miRNeasy Mini Kit (Cat.no 217004, Qiagen,). Approximately 10ug of total RNA was obtained per genotype that was send to Fasteris (Fasteris SA, Geneva, CH) for transcriptome sequencing and bioinformatic analysis. The HiSeq run was performed on a Hi-Seq 2000 with 1X100+7 number of sequencing cycles using the TruSeq SBS v5 with the data analysis pipeline carried out by HiSeq Control Soft. V. 1.1.37.8, RTA 1.7.48, CASAVA 1.7. Sequences were aligned to the Drosophila melanogaster genome revision 5.30. Counts were normalized as reads per million (RPM) by dividing by the total number of reads and multiplying by 1 million.

Data Analysis

For each gene I determined the presence of a signal sequence using SignalP [12]. Expression patterns, including male biased expression, outside the accessory gland were obtained by analyzing data reported by Flybase [42], Fly Atlas [10], and ModENCODE [14,15,43]. For genes with no listed function in Flybase, I used Pfam [29]to identify conserved domains where possible.

Fly stocks and Media

All flies were raised at room temperature (23±1°C) in glass bottles on standard yeast-glucose media (See previous Chapters). Females were aged 3-5 days from eclosion in groups of 5-12 in

glass vials with added yeast. Male flies were aged 3-5 days from eclosion in groups of 10-20 in glass vials on standard yeast-glucose media. Fly lines containing a UAS-hairpin specific to each of my genes of interest were obtained from the Vienna Drosophila RNAi Center (VDRC) [9]. To generate knockdown males, each UAS-hairpin line was crossed to *iab-6D1*-GAL4/CyO; control flies were generated by crossing the driver to *,w[1118];P{attP,y[+],w[3`] (Attp*^{60100}) in the case of KK insertion lines or w^{1118} in the case of GD insertion lines. Balancer sibling controls were used where appropriate. Knockdown of transcripts relative to controls were confirmed by RT-PCR as described in Chapter 2-4 except that mRNA samples were extracted from dissected accessory glands rather than whole flies [16]. To generate GFP expression in the secondary cells, the *Abd-B* reporter BAC (see Chapter 5) was integrated into the *Dicer*;*iab-6D5*-GAL4/CyO driver line and crossed to RNAi lines as previously mentioned.

Fertility/fecundity assays (FFA)

I performed assays as described in Chapters 2-5) using 3-5 day old Canton-S virgin females. Single females were placed in glass vials with food. Under observation so that mating could be confirmed, each female was allowed access to an RNAi male or control male. After mating finished, the male was removed. Individual females were transferred to fresh food every 24 hours for 10 days and the eggs laid in the previous vial were counted as described in [3] (See Chapter 5). Wilcoxon non-parametric test were used to compare mates of RNAi and Control males in total and on individual days. The overall 10 day trends were analyzed by rmANOVA. All statistical analysis was performed with the JMP9 software [44]. Upon eclosion, all progeny from each vial were counted. Hatchability (# progeny / # eggs) was calculated per day and across the 10-day period for each female.

Receptivity assays

Matings were performed as described above for FFAs. Single mated females were kept in a vial on yeast-glucose media for 4 days after the start of mating (ASM). Females that did not produce viable progeny during the 4d holding period were discarded from the assay to ensure that only healthy females that successfully mated in the initial were included in my analysis. On day 4, each female was moved to a fresh vial and allowed access to a single Canton S male. Couples were observed for an hour at 15 min intervals, after which the proportion of successful rematings was recorded.. Comparisons of remating frequency between females mated to either control or RNAi males were conducted using a Wilcoxon ranked sums test (WRST) using JMP9 software [44].

Western Blots

Matings were carried out as described in Chapter 5 [3].. All samples were frozen in liquid nitrogen before being stored at -20°C until dissection. Protein samples and Western blot analyses were performed as in [3,6,11].

PNGase F assays were performed with reagents from New England Biolabs Inc. and as outlined in Chapter 5 except that only 8 male accessory glands from control or RNAi males were used in the initial dissections.

6.5 REFERENCES

1. Styger D (1992) Molekulare Analyse des Sexpeptidgens aus Drosophila melanogaster. PhD Thesis University of Zurich: Zurich, Switzerland.

2. Liu H, Kubli E (2003) Sex-peptide is the molecular basis of the sperm effect in Drosophila melanogaster. Proc Natl Acad Sci U S A 100: 9929-9933.

3. Gligorov D, Sitnik JL, Maeda RK, Wolfner MF, Karch F (2013) A novel function for the Hox gene Abd-B in the male accessory gland regulates the long-term female post-mating response in Drosophila. PLoS Genet 9: e1003395.

4. Minami R, Wakabayashi M, Sugimori S, Taniguchi K, Kokuryo A, et al. (2012) The homeodomain protein defective proventriculus is essential for male accessory gland development to enhance fecundity in Drosophila. PLoS One 7: e32302.

5. Leiblich A, Marsden L, Gandy C, Corrigan L, Jenkins R, et al. (2012) Bone morphogenetic protein- and mating-dependent secretory cell growth and migration in the Drosophila accessory gland. Proc Natl Acad Sci U S A 109: 19292-19297.

6. Ram KR, Wolfner MF (2009) A network of interactions among seminal proteins underlies the long-term postmating response in Drosophila. Proc Natl Acad Sci U S A 106: 15384-15389.

7. Chapman T, Herndon LA, Heifetz Y, Partridge L, Wolfner MF (2001) The Acp26Aa seminal fluid protein is a modulator of early egg hatchability in Drosophila melanogaster. Proc Biol Sci 268: 1647-1654.

8. LaFlamme BA, Ram KR, Wolfner MF (2012) The Drosophila melanogaster seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. PLoS Genet 8: e1002435.

9. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, et al. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448: 151-156.

10. Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39: 715-720.

11. Ravi Ram K, Ji S, Wolfner MF (2005) Fates and targets of male accessory gland proteins in mated female Drosophila melanogaster. Insect Biochem Mol Biol 35: 1059-1071.

12. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785-786.

13. (1994) FlyBase--the Drosophila database. The FlyBase Consortium. Nucleic Acids Res 22: 3456-3458.

14. Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, et al. (2010) Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science 330: 1787-1797.

15. McQuilton P, St Pierre SE, Thurmond J (2012) FlyBase 101--the basics of navigating FlyBase. Nucleic Acids Res 40: D706-714.

16. Ram KR, Wolfner MF (2007) Sustained post-mating response in Drosophila melanogaster requires multiple seminal fluid proteins. PLoS Genet 3: e238.

17. Findlay GD, MacCoss MJ, Swanson WJ (2009) Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. Genome Res 19: 886-896.

18. Walker DL, Wang D, Jin Y, Rath U, Wang Y, et al. (2000) Skeletor, a novel chromosomal protein that redistributes during mitosis provides evidence for the formation of a spindle matrix. J Cell Biol 151: 1401-1412.

19. Rath U, Wang D, Ding Y, Xu YZ, Qi H, et al. (2004) Chromator, a novel and essential chromodomain protein interacts directly with the putative spindle matrix protein skeletor. J Cell Biochem 93: 1033-1047.

20. Fabian L, Xia X, Venkitaramani DV, Johansen KM, Johansen J, et al. (2007) Titin in insect spermatocyte spindle fibers associates with microtubules, actin, myosin and the matrix proteins skeletor, megator and chromator. J Cell Sci 120: 2190-2204.

21. Ding Y, Yao C, Lince-Faria M, Rath U, Cai W, et al. (2009) Chromator is required for proper microtubule spindle formation and mitosis in Drosophila. Dev Biol 334: 253-263.

22. Yao C, Ding Y, Cai W, Wang C, Girton J, et al. (2012) The chromodomain-containing NH(2)-terminus of Chromator interacts with histone H1 and is required for correct targeting to chromatin. Chromosoma 121: 209-220.

23. Qi H, Rath U, Wang D, Xu YZ, Ding Y, et al. (2004) Megator, an essential coiled-coil protein that localizes to the putative spindle matrix during mitosis in Drosophila. Mol Biol Cell 15: 4854-4865.

24. Vaquerizas JM, Suyama R, Kind J, Miura K, Luscombe NM, et al. (2010) Nuclear pore proteins nup153 and megator define transcriptionally active regions in the Drosophila genome. PLoS Genet 6: e1000846.

25. Bairati A (1968) Structure and ultrastructure of the male reproductive system in *Drosophila melanogaster*. Monitore zoologico italiano 2: 105-182.

26. Zheng Y (2010) A membranous spindle matrix orchestrates cell division. Nat Rev Mol Cell Biol 11: 529-535.

27. Bairati A (1966) Filamentous structures in spermatic fluid of Drosophila melanogaster. J Microscopie 5: 265.

28. Chen PS (1984) The Functional Morphology and Biochemistry of Insect Male Accessory Glands and Thier Secretions Annu Rev Entomol 29: 233-255.

29. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: D290-301.

30. Chen PS, Stumm-Zollinger E, Aigaki T, Balmer J, Bienz M, et al. (1988) A male accessory gland peptide that regulates reproductive behavior of female D. melanogaster. Cell 54: 291-298.

31. Findlay GD, Yi XH, MacCoss MJ, Swanson WJ (2008) Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. Plos Biology 6: 1417-1426.

32. Tandler B, Williamson DL, Ehrman L (1968) Unusual filamentous structures in the paragonia of male Drosophila. J Cell Biol 38: 329-336.

33. Peng J, Chen S, Busser S, Liu HF, Honegger T, et al. (2005) Gradual release of sperm bound sex-peptide controls female postmating behavior in Drosophila. Current Biology 15: 207-213.

34. Hinton BT, Howards SS (1982) Rat testis and epididymis can transport [3H] 3-O-methyl-D-glucose, [3H] inositol and [3H] alpha-aminoisobutyric acid across its epithelia in vivo. Biol Reprod 27: 1181-1189.

35. Gonzales GF (2001) Function of seminal vesicles and their role on male fertility. Asian J Androl 3: 251-258.

36. Caballero Peregrin B, Gonzalez Guerbolini C, Farinas Fernandez F (1979) [Fructose concentration in sperm liquid and its correspondence with sperm motility: vertical progression of the spermatozoons and the Botella-Casares index]. Acta Ginecol (Madr) 33: 367-372.

37. Vivas-Acevedo G, Lozano-Hernandez R, Camejo MI (2011) Markers of accessory sex glands function in men with varicocele, relationship with seminal parameters. Can J Urol 18: 5884-5889.

38. Gonzales GF, Villena A (2001) True corrected seminal fructose level: a better marker of the function of seminal vesicles in infertile men. Int J Androl 24: 255-260.

39. King M, Eubel H, Millar AH, Baer B (2011) Proteins within the seminal fluid are crucial to keep sperm viable in the honeybee Apis mellifera. J Insect Physiol 57: 409-414.

40. Poole HK, Edwards JF (1970) Induction of motility in honey bee (Apis mellifera L.) spermatozoa by sugars. Experientia 26: 859-860.

41. Collins AM, Caperna TJ, Williams V, Garrett WM, Evans JD (2006) Proteomic analyses of male contributions to honey bee sperm storage and mating. Insect Mol Biol 15: 541-549.

42. (1999) The FlyBase database of the Drosophila Genome Projects and community literature. The FlyBase Consortium. Nucleic Acids Res 27: 85-88.

43. Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, et al. (2012) Identification and properties of 1,119 candidate lincRNA loci in the Drosophila melanogaster genome. Genome Biol Evol 4: 427-442.

44. (1989-2007) JMP, Version 9. Cary, NC: SAS Institute Inc.

CHAPTER 7

GENERAL DISCUSSION

In this thesis I have presented several studies that identified and characterized new Sfps and female proteins necessary for regulating the PMR. First, I showed that *Drosophila Neprilysins* play a role in regulating both male and female fertility and that *Nep2* in particular is a good candidate for regulation by Sfps, in females. Second, I demonstrated the functions of three gene duplicates (CG32834 and CG9897 in the female and CG32833 in the male) in regulating the female PMR and showed that CG32833 is necessary for long term egg-laying. Third, I showed that Sfp intrepid is an LTR protein and attempted to place it into the LTR network. Fourth, I demonstrated that the secondary cells of the accessory gland are essential for regulating the LTR. Lastly, through examining the contribution of individual genes impacted by the *iab-6*^{cocu} mutation I identified 8 genes, at least one of which encodes a transferred Sfp, whose expression in the secondary cells is essential for the maintenance of the LTR. The impact of my work on our understanding of the LTR is summarized in Figures 7.1 and 7.2.

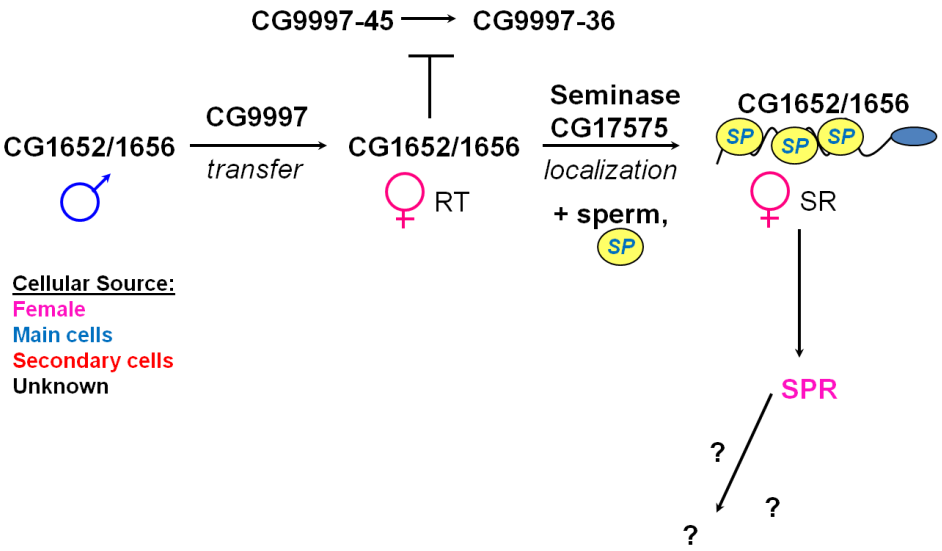
The secondary cells are essential for regulating the LTR

Prior to this work, little was known about the function of the secondary cells in regulating the PMR much less the products that they produce. We now know that the secondary cells make

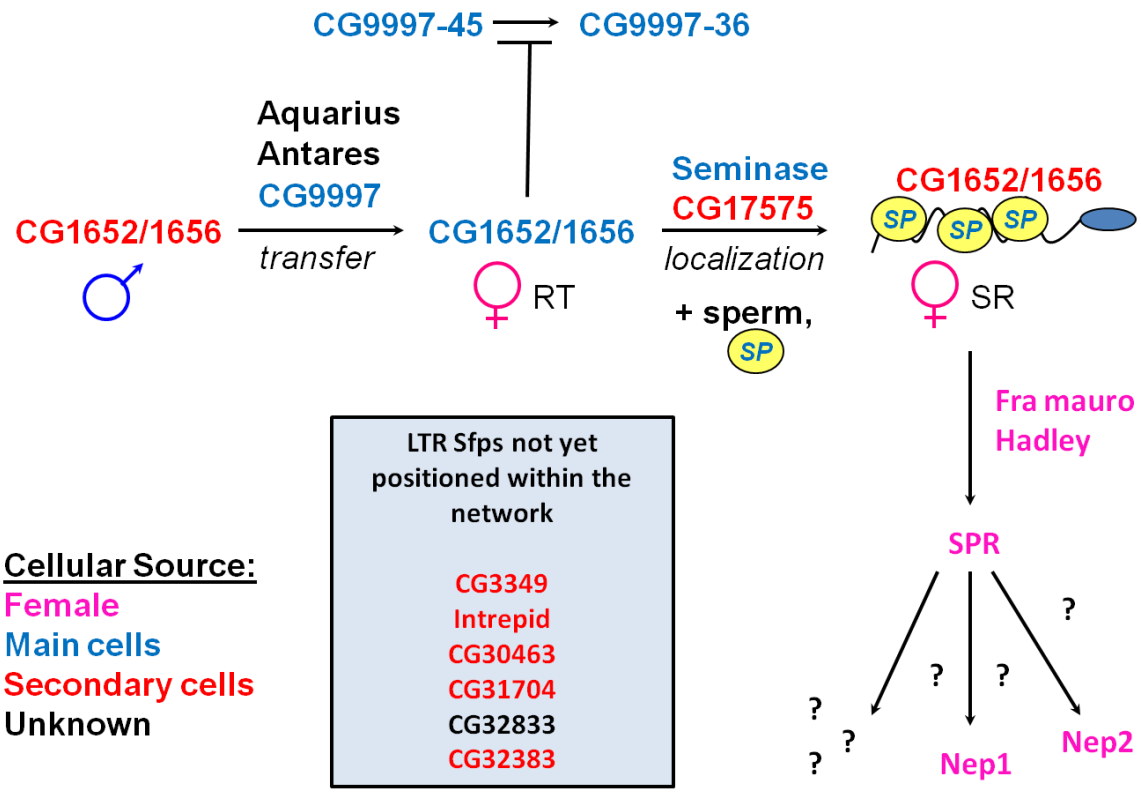
Figure 7.1: Summary of the LTR Network

A) The LTR network consists of 5 Sfps that are required for SP to be stored normally. In the female, SP interacts with SPR to influence the PMR. Only the cellular origin of SP was previously known. B) 8 new Sfps and 2 female proteins were identified as being essential for the LTR. Two of these Sfp, Aquarius and Antares, were able to be placed in the LTR network at the same step as CG9997, at transfer of CG1656 and CG1652. The female proteins Fra mauro and Hadley do not impact SP storage and are likely either essential for SP to interact with SPR or are downstream of SPR. The remaining 6 Sfps have not yet been placed in the network. We identified 2 possible downstream targets of the LTR network, Nep2 and Nep1, which are both essential for normal female egg-laying. Further, the cellular source of all but 3 of the Sfps in the LTR Network has now been identified.

A) The Original LTR Network[1]



B) The Extended LTR Network



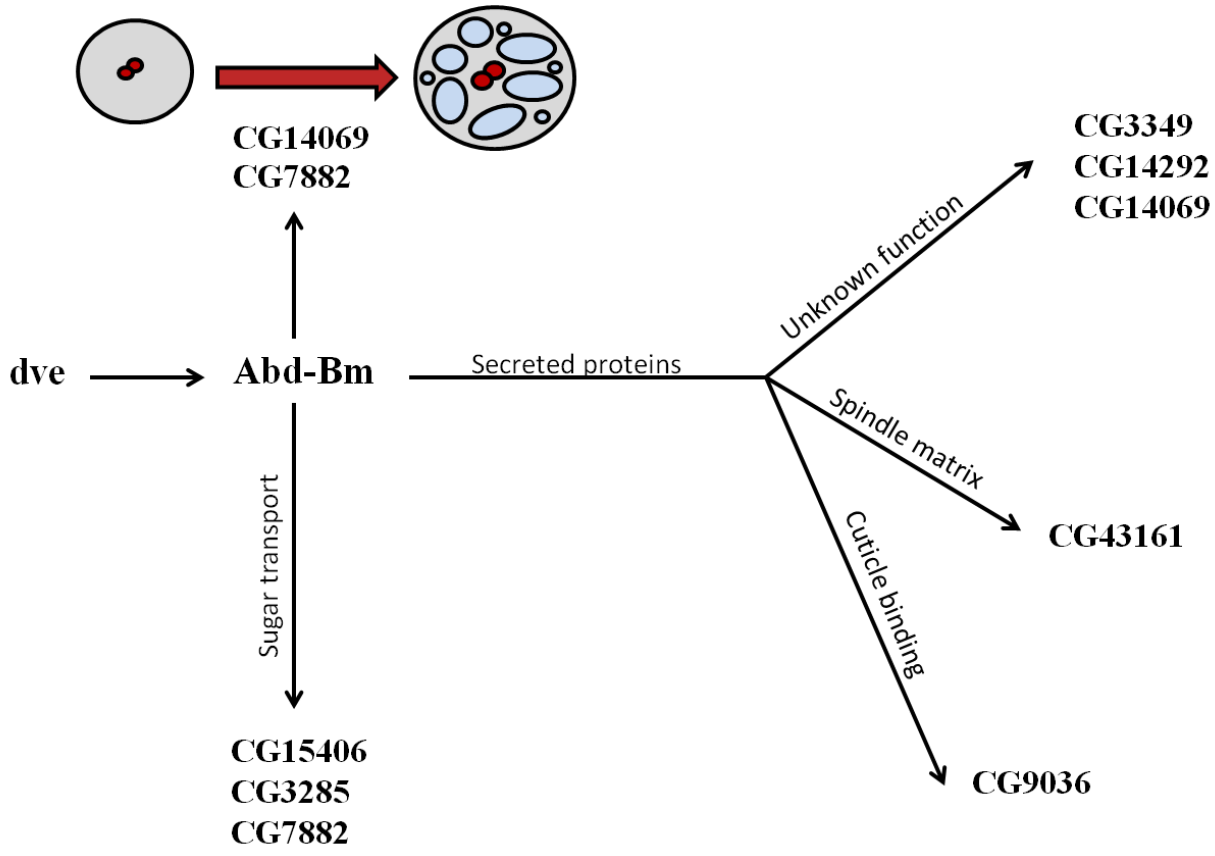


Figure 7.2: Summary of *Abd-B* functions in the LTR

I identified 8 genes down regulated in the secondary cells of *Abd-B* mutant males whose products are essential for the LTR. Two of these genes also influence vacuole formation in the secondary cells. While 5 of these genes encode a signal sequence suggesting that they are secreted proteins that could be transferred to females during mating , only one of these proteins (CG3349) is a known Sfp. In the graphic above, these genes are delineated by their predicted function. These results suggest that *Abd-B* influences multiple cellular functions (particularly in cellular transport) in addition to regulating the expression of Sfps.

at least 9 Sfps (Chapter 3, 5, 6, Appendix C) that are essential for regulating the LTR. Further, the off-target gene CG31326 impacts 24h egg-laying when knocked down in the secondary cells, suggesting that these cells may also be important for the STR. In addition to individual Sfps, it is clear that a wide array of systems in the secondary cell are necessary for the LTR, including sugar transport, spindle matrix, and cuticle binding proteins. This study only investigates a handful of genes influenced by *Abd-B* expression. More than a hundred untested genes are also differentially expressed in *iab-6*^{cocu} mutants, including several up regulated Sfps. The function of these genes in regulating the PMR is yet to be discovered and holds promise not only for identifying new Sfps but also for understanding the cellular functions of the secondary cells.

Similarly, while our understanding of the Sfp contribution of the secondary cells has vastly improved through the work I present here, this is likely not a complete list of secondary cell Sfps. My work suggests that the *iab-6*^{cocu} mutation only impacts the expression of a handful (14) known Sfps. Further, 3 of the now known secondary cell Sfps, CG1656, CG1652, and CG17575, are not affected by the loss of *Abd-B* expression in these cells. The most direct way of studying the total suite of proteins produced by the secondary cells is through targeted cell ablation. To do this the *iab-6*D1-GAL4 driver can be used to drive expression of either UAS-*reaper* to kill the cell or diphtheria toxin to prevent protein translation. To get around lethality caused by nonspecific expression of this driver in the developing CNS, we can take advantage of a temperature sensitive version of the GAL4 suppressor GAL80 [2]. Preliminary work using this system has shown promise for ablating the secondary cells (data not shown), however further work is needed to optimize this system and identify the cellular source of the remaining Sfps.

In a similar vein, the morphology data collected by my collaborators D. Gligorov and F. Karch, suggest that the vacuoles of the secondary cells may be important for their function in

regulating the LTR. Although there is currently no evidence that Sfps required for the LTR are present in the vacuole, other proteins including the Sfp Ovulin [3] and Angiotensin converting enzyme (ACE)[4] are detected within these structures. Further work is needed to elucidate the contents of the vacuoles to determine how they may influence Sfps.

Perhaps the most necessary next step is placing the remaining Sfps into the LTR Network, similar to what we did in Chapter 4. However, the playing field has increased and more antibodies are needed to increase the resolution of the network. In addition, while we have identified new LTR Network proteins in both males and females, there is still no evidence that these proteins physically interact with one another. Experiments using Co-IP and/or Y-2H to identify interactions were not successful previously, possibly because interactions between the original 5 LTR network proteins were not direct or part of a complex. The expansion of interactors and the ability to guess at interactions using ERC should help increase the likelihood of success.

In conclusion, I have demonstrated that the secondary cells work together with main cells to regulate the LTR in *Drosophila melanogaster*. I have also identified new proteins in both males and females that contribute to the regulation of the PMR. This work opens up new possibilities for how Sfps interact with one another and the female as well as how Sfps are regulated prior to transfer.

REFERENCES

1. Ram KR, Wolfner MF (2009) A network of interactions among seminal proteins underlies the long-term postmating response in Drosophila. Proc Natl Acad Sci U S A 106: 15384-15389.

2. Bohm RA, Welch WP, Goodnight LK, Cox LW, Henry LG, et al. (2010) A genetic mosaic approach for neural circuit mapping in Drosophila. Proc Natl Acad Sci U S A 107: 16378-16383.

3. Monsma SA, Harada HA, Wolfner MF (1990) Synthesis of two Drosophila male accessory gland proteins and their fate after transfer to the female during mating. Dev Biol 142: 465-475.

4. Rylett CM, Walker MJ, Howell GJ, Shirras AD, Isaac RE (2007) Male accessory glands of Drosophila melanogaster make a secreted angiotensin I-converting enzyme (ANCE), suggesting a role for the peptide-processing enzyme in seminal fluid. J Exp Biol 210: 3601-3606.

APPENDIX A

MOLECULAR SOCIAL INTERACTIONS: *DROSOPHILA MELANOGASTER* SEMINAL FLUID PROTEINS AS A CASE STUDY[a]

While studies of social behavior generally focus on observable interactions between individuals, additional "hidden" social interactions occur on the molecular level. These molecular interactions can be considered social in two ways. First, observable social interactions are influenced by molecular interactions (Ellison and Gray, 2009). Second, molecules from different individuals can interact in what we call here "molecular social interactions." The molecular biology of social behavior has thus far been focused primarily on the former: molecular interactions within an animal that either induce or result from social interactions. This approach has successfully identified molecular interactors in rodent and avian affiliative behavior (e.g., reviewed in Adkins-Regan, 2009 and Keverne and Curley, 2004), nematode feeding behavior (e.g., reviewed in de Bono and Maricq, 2005), eusocial behavior (e.g., Smith *et al.*, 2008), and Drosophila courtship (e.g., reviewed inDickson, 2008 and Villella and Hall, 2008). However, a complete molecular understanding of social behavior necessitates an understanding not just of how molecules interact within a social animal, but also how "social molecules" interact among animals. Here, we present a case study of such "molecular social interactions" that involves *Drosophila melanogaster* seminal fluid proteins (Sfps) that are

[a] This is an excerpt of my contribution to the review mentioned in Chapter 1:

Sirot LK, LaFlamme BA, Sitnik JL, Rubinstein CD, Avila FW, et al. (2009) Molecular social interactions: Drosophila melanogaster seminal fluid proteins as a case study. Adv Genet 68: 23-56.

Reprinted with permission.

produced in the male reproductive tract, and transferred to the female along with sperm during mating. In the case of *D. melanogaster* Sfps, the molecular social interactions are extensive, as gene products in seminal fluid induce short- and long-term changes in females' behavior, physiology, and gene expression, and these changes require interactions of Sfps with female-derived molecules and physiology (e.g., muscle, circulatory, and neural systems). Thus, the male- and female-derived molecules are involved in an inherently social interaction—that is an interaction between two individuals of the same species. Molecular social interactions affect the outcome of individual matings and occur directly between males and between males and females, and indirectly between multiple males that have mated with a given female. As we will discuss, molecular social interactions both shape, and are shaped by, observable behavioral interactions between conspecifics to affect lifetime reproductive success.

Following mating, female *D. melanogaster* display a number of behavioral and physiological changes that impact both male and female reproductive success. For example, after mating, females increase their rates of oogenesis, ovulation, egg-laying, and food intake (e.g., reviewed in Chapman, 2001, Chapman and Davies, 2004 and Wolfner, 2009). Sperm from the male are stored in specialized sperm storage organs (Fig. 2.1), and this process may be facilitated by changes in uterine shape beginning at the onset of mating (Adams and Wolfner, 2007 and Avila and Wolfner, 2009). For several days, mated females are less likely to accept suitors, actively fleeing or kicking any persistent male (Ringo, 1996 and Spieth and Ringo, 1983). Within hours after mating, the female increases expression of several known antimicrobial peptide genes (Kapelnikov et al., 2008a, Lawniczak and Begun, 2004, Mack et al., 2006, McGraw et al., 2004 and Peng et al., 2005b), yet the realized immune response that protects the female from infections is reduced (Fedorka *et al.*, 2007). The lifespan

of *Drosophila* females is also reduced by mating (Barnes et al., 2008, Chapman et al., 1995, Civetta and Clark, 2000b, Fowler and Partridge, 1989 and Wigby and Chapman, 2005).

These changes in behavior, physiology, and gene expression may be brought about by the behavioral act of mating, by the transfer of sperm, or by other contents of the seminal fluid. Since males that do not produce sperm still elicit postmating responses in their partners (albeit, weaker and/or more short term; Kalb et al., 1993, Manning, 1962, Manning, 1967 and Xue and Noll, 2000), nonsperm components of the seminal fluid must be involved in the induction of these responses. In fact, males that transfer sperm but do not transfer Sfps produced in their accessory glands (Fig. 2.1) fail to elicit most postmating responses in females (Kalb et al., 1993 and Xue and Noll, 2000). It is known that the ejaculatory duct and ejaculatory bulb also produce secreted proteins that constitute part of seminal fluid, and that some of these proteins are necessary for postmating responses (e.g., Gilbert et al., 1981, Iida and Cavener, 2004, Ludwig et al., 1991, Lung and Wolfner, 2001, Lung et al., 2001, Meikle et al., 1990, Samakovlis et al., 1991 and Bretman et al., in press. These results together demonstrate that sperm and Sfps are both required to induce long-term postmating responses in females (Heifetz et al., 2001, Kalb et al., 1993, Kubli, 2003, Manning, 1962 and Manning, 1967).

Sfps comprise an elaborate intraspecific signaling system. Of the more than 180 predicted extracellular proteins present in the reproductive secretory glands of male *D. melanogaster*, over 100 have been confirmed to be transferred to the female along with sperm (e.g., reviewed in Ravi Ram and Wolfner, 2007a and Chapman, 2008; see also Chintapalli et al., 2007, Findlay et al., 2008, Findlay et al., 2009,Takemori and Yamamoto, 2009 and Walker et al., 2006). Many of the transferred proteins fall into conserved protein classes found in the seminal fluid of most animals

studied to date and include proteases, protease inhibitors, acid lipases, cysteine-rich secretory proteins (CRISPs), and lectins (Mueller et al., 2004 and Ravi Ram and Wolfner, 2007a). Other, less-expected, classes of Sfps such as odorant-binding proteins suggest a possible role for small molecules in inducing female postmating responses (Findlay *et al.*, 2008). Odorant-binding proteins are known to shuttle pheromones or other small molecules to odorant receptors in the olfactory system (e.g., reviewed in Pelosi *et al.*, 2005). Presence of predicted odorant-binding proteins in the seminal fluids suggests that they may play a similar shuttling role for molecules once they are within the female reproductive tract. The wide variety of protein classes present in the seminal fluid suggests that Sfps take part in a complex series of interactions within the mated female and do not just fulfill a single simple role.

Upon transfer to females, Sfps target to specific tissues which are likely to relate to their function within the mated female (e.g., Bertram et al., 1996, Heifetz et al., 2000, Lung and Wolfner, 1999, Meikle et al., 1990,Peng et al., 2005a and Ravi Ram et al., 2005; Fig. 2.1). For example, proteins associated with sperm storage and retention have been detected in the female sperm storage organs, and ovulin, which stimulates ovulation, targets to the base of the ovaries (Heifetz et al., 2000 and Ravi Ram et al., 2005). Several Sfps, including ovulin, have also been detected in the circulatory system of mated females from where they can gain access to the brain and/or endocrine systems (Lung and Wolfner, 1999, Meikle et al., 1990, Pilpel et al., 2008 and Ravi Ram et al., 2005) and thus, potentially, affect female behavior. Further studies of the targets of Sfps may help to uncover their functions in the mated female.

D. melanogaster Sfps provide an excellent model system in which to investigate molecular social interactions, due to the powerful tools available in this species. Mutant or transgenic males in which Sfps are increased, decreased, or eliminated can be used to dissect the effect(s) of

particular Sfps on female postmating responses (e.g., Bretman et al., in press, Chapman et al., 2003, Gilbert et al., 1981, Herndon and Wolfner, 1995, Iida and Cavener, 2004, Liu and Kubli, 2003, Mueller et al., 2008, Neubaum and Wolfner, 1999, Ravi Ram and Wolfner, 2007b, Ravi Ram et al., 2006 and Wong et al., 2008a). A large collection of freely available genomic databases (e.g., FlyBase; FlyAtlas, Chintapalli *et al.,* 2007) facilitate rapid progress as well. These techniques and tools, along with studies associating allelic variation in Sfps with variation in their effects, have led to a greater understanding of the molecular social interactions taking place between all of the players involved in *Drosophila* mating (e.g., reviewed in Wolfner, 2009). Furthermore, studies of *D. melanogaster* Sfps are likely to provide insights into the molecular social interactions of other species given that Sfps impact female postmating responses across a wide taxonomic range (e.g., reviewed in Gillott, 2003 and Poiani, 2006). We will use two particularly well-studied Sfps, the sex peptide (SP) and ovulin, as examples in the following sections to illustrate the way in which Sfps act as molecular mediators for social interactions. SP is a small peptide that affects female response to male courtship, oogenesis, and her ovulation, immune response, feeding, and juvenile hormone production (Carvalho et al., 2006, Chapman et al., 2003, Domanitskaya et al., 2007, Kubli, 2003, Liu and Kubli, 2003 and Moshitzky et al., 1996). Ovulin is a large prohormone that increases ovulation during the first 24 h after mating. Further details of both these proteins, as well as the social context in which they exert their functions, are discussed herein.

While over 180 known or putative *D. melanogaster* Sfps have been identified, only one female receptor to an Sfp is known: the SP receptor (SPR), a G-protein-coupled receptor expressed in the female reproductive tract and nervous system (Yapici *et al.,* 2008). However, we expect that many Sfps interact with female-derived proteins. Some female-derived proteins

that play a role in female postmating behavior and physiology have been identified and will be discussed in this review, but their interactions with Sfps remain speculative at this time. Several approaches have been used to identify genes in females whose products mediate response to, are regulated by, or otherwise interact with, Sfps. Proteins produced in the female sperm storage organs have been identified and have the potential to interact with Sfps (Allen and Spradling, 2008, Lawniczak and Begun, 2007, Prokupek et al., 2008 and Prokupek et al., 2009). Microarray data from whole flies, heads, or reproductive tract tissues have shown that different aspects of mating, including Sfps, cause a transcriptional response in the female after mating (Innocenti and Morrow, 2009, Kapelnikov et al., 2008a, Lawniczak and Begun, 2004, Lawniczak and Begun, 2007, Mack et al., 2006, McGraw et al., 2004, McGraw et al., 2008, McGraw et al., 2009 and Peng et al., 2005b), though it is not likely that most initial postmating responses are due to mating-induced transcription. Transcriptional changes of the largest magnitude are seen by about 6–8 h after mating, a time by which most Sfps are no longer detectable in the female. Therefore, Sfps may set into motion the transcriptional modification of the female, but the genes regulated by these modifications are less likely to encode Sfp-interacting proteins than the genes expressed by the female prior to mating. Nevertheless, these mating-regulated genes likely are players in the next steps of the molecular social interactions. To fully understand the molecular social interactions in which Sfps are involved, we must identify female interactors, their functions, and how they have coevolved with their male-derived partners.

APPENDIX B

TABLE B: LIST OF ALL CURRENTLY KNOWN SFPS

CG	LOCATION	NAME	SYMBOL	FUNCTION	TRANSFER	LIST OVERLAP
CG10000	3R	-	CG10000	lectin		Wolfner
CG10029	3R	-	CG10029	protein disulfide isomerase		Takemori
CG10041	3R	-	CG10041	serine protease	yes	Findlay
CG10112	2R	Cuticular protein 51A	Cpr51A	peptide/Prohormone	yes	Findlay
CG10184	3R	-	CG10184	threonine adolase		Takemori
CG10284	3R	-	CG42564	CRISP	yes	Findlay, Wolfner
CG10363	2L	Thioester-containing protein 4	Tep4	alpha-macroglobulin		Takemori, Wolfner
CG10407	3R	-	CG10407	Tyrosine protein kinase	yes	Findlay
CG10433	2R	-	CG10433	defensin	yes	Findlay, Wolfner
CG10576	3L	-	CG10576	Metallopeptidase		Takemori
CG10586	3L	Seminase	Sems	serine protease	yes	Findlay, Wolfner
CG10587	3L	-	CG10587	serine protease	yes	Findlay, Wolfner
CG10651	2L	-	CG10651	CRISP	yes	Findlay, Wolfner
CG10688	3L	-	CG10688	phosphomannomutase		Takemori
CG10730	2L	-	CG10730	Alkaline-phosphatase	yes	Findlay
CG10852	3L	Accessory gland protein 63F	Acp63F	peptide/Prohormone	yes	Findlay, Takemori, Wolfner
CG10862	3L	-	CG10862	ubiquitin-protein ligase activity	yes	Findlay
CG10956	2R	Serpin 53F	Spn53F	serpin		Wolfner
CG11037	3L	-	CG11037	serine protease	yes	Findlay, Wolfner
CG11112	2R	-	CG11112	peptide/Prohormone	Simulins	Findlay, Takemori, Wolfner
CG11113	2R	-	CG11113	peptide/Prohormone		Wolfner
CG11131	3L	-	CG11131	peptide/Prohormone		Wolfner
CG1152	3R	Glucose dehydrogenase	Gld	Glucose dehydrogenase		Wolfner
CG11598	3R	-	CG11598	acid lipase	yes	Findlay, Wolfner
CG11608	3R	-	CG11608	Triglycerol lipase	yes	Findlay, Wolfner
CG11630	2L	-	CG11630	Alkaline-phosphatase	yes	Findlay
CG11664	X	-	CG11664	serine protease		Wolfner
CG11828	3R	-	CG11828	Prolyl 4-Hydroxylase		Takemori
CG11864	2L	-	CG11864	metalloprotease	yes	Findlay, Wolfner
CG11977	3R	-	CG11977	CRISP		Wolfner
CG12030	3L	UDP-galactose 4'-epimerase	Gale	Epimerase		Takemori
CG12233	X	lethal (1) G0156	l(1)G0156	isopropylmalate dehydrogenase		Takemori
CG12558	3R	Intrepid	CG12558	serine protease		Wolfner
CG1262	3L	Accessory gland protein 62F	Acp62F	trypsin protease inhibitor	yes	Findlay, Wolfner
CG12879	3R	-	CG43320	cupin	yes	Findlay, Takemori

CG	LOCATION	NAME	SYMBOL	FUNCTION	TRANSFER	LIST OVERLAP
CG12943	2R	-	CG12943	transmembrane transporter	yes	Findlay
CG1319	3L	-	CG1319	electron carrier activity	yes	Findlay
CG13309	3L	-	CG13309	peptide/Prohormone		Wolfner
CG13340	2R	Sperm-Leucylaminopeptidase 7	S-Lap7	Leucyl aminopeptidase	yes	Findlay
CG1361	3R	Andropin	Anp	Cecropin	yes	Findlay
CG13759	X	-	CG13759	acetyltransferase		Takemori
CG13873	2R	Odorant-binding protein 56g	Obp56g	Oderant Binding	yes	Findlay
CG13965	2L	-	CG13965	peptide/Prohormone	yes	Findlay, Wolfner
CG14034	2L	-	CG14034	phospholipase	yes	Findlay, Wolfner
CG14061	3R	Aquarius	CG14061	serine protease	yes	Findlay, Takemori, Wolfner
CG14476	X	-	CG14476	Glycosyl hydrolases		Takemori
CG14560	3L	male-specific opa containing gene	msopa	peptide/Prohormone	yes	Takemori, Wolfner
CG1462	3R	Alkaline phosphatase 4	Aph-4	alkaline phosphatase		Wolfner
CG14748	2R	-	CG42326	peptide/Prohormone		Takemori, Wolfner
CG14913	2L	-	CG14913	peptide/Prohormone		Wolfner
CG14996	3L	Chd64	Chd64	Calponin		Takemori
CG15031	X	Protein phosphatase Y regulator 1	PPYR1	unknown	yes	Findlay
CG15116	2R	-	CG15116	Glutathione Peroxidase	yes	Findlay, Wolfner
CG15117	2R	-	CG15117	Glycosyl hydrolases	yes	Findlay
CG15616	2R	Accessory gland protein 53C14b	Acp53C14b	peptide/Prohormone	yes	Findlay, Takemori, Wolfner
CG15635	2L	-	CG15635	unknown	yes	Findlay, Takemori
CG15641	X	-	CG15641	peptide/Prohormone	yes	Findlay, Wolfner
CG15841	2L	-	CG15841	peptide/Prohormone		Wolfner
CG1633	X	thioredoxin peroxidase 1	Jafrac1	thioredoxin peroxidase		Takemori
CG1652	2R	lectin-46Cb	lectin-46Cb	C-type lectin	yes	Findlay, Wolfner
CG1656	2R	lectin-46Ca	lectin-46Ca	C-type lectin	yes	Findlay, Takemori, Wolfner
CG16707	3L	visgun	vsg	peptide/Prohormone		Takemori, Wolfner
CG16995	2L	-	CG16995	CRISP		Wolfner
CG1701	2R	-	CG1701	unknown	yes	Findlay, Takemori
CG17011	2L	lectin-30A	lectin-30A	lectin	yes	Findlay, Wolfner
CG17050	2R	-	CG43691	unknown	yes	Findlay
CG17058	X	Peritrophin A	Peritrophin-A	Peritrophin-A	yes	Findlay
CG17097	2L	-	CG17097	acid lipase	yes	Findlay, Wolfner

CG	LOCATION	NAME	SYMBOL	FUNCTION	TRANSFER	LIST OVERLAP
CG17137	2L	Porin2	Porin2	porin	yes	Findlay, Takemori
CG1721	3R	Phosphoglycer omutase	Pglym78	Phosphoglyceromuta se		Takemori
CG17242	2L	-	CG17242	serine protease	yes	Findlay
CG17271	3R	-	CG17271	peptide/Prohormone		Wolfner
CG17472	2L	-	CG17472	peptide/Prohormone	yes	Findlay, Wolfner
CG17575	2R	-	CG17575	CRISP	yes	Findlay, Takemori, Wolfner
CG17673	3L	Accessory gland protein 70A	Acp70A	peptide/Prohormone	yes	Findlay, Takemori, Wolfner
CG17797	2L	Accessory gland protein 29AB	Acp29AB	C-type lectin	yes	Findlay, Wolfner
CG17799	2L	lectin-29Ca	lectin-29Ca	lectin	yes	Findlay, Takemori, Wolfner
CG17843	3R	-	CG17843	thioredoxin	yes	Findlay, Wolfner, Takemori
CG1787	X	Hexosaminidas e 2	Hexo2	Hexosaminidase	yes	Findlay, Takemori
CG17919	3R	-	CG17919	anion binding	yes	Findlay
CG17924	3R	Accessory gland protein 95EF	Acp95EF	peptide/Prohormone	yes	Findlay, Wolfner
CG1803	X	regucalcin	regucalcin	unknown	yes	Findlay
CG18067	2R	-	CG18067	unknown	yes	Findlay, Takemori
CG18135	3L	-	CG18135	phosphoddiesterase	yes	Findlay
CG18233	3L	-	CG18233	Oxidoreductase		Wolfner
CG18234	3L	-	CG18234	Oxidoreductase		Wolfner
CG18284	2L	-	CG18284	acid lipase	yes	Findlay, Wolfner
CG18348	3L	Cuticular protein 67Fb	Cpr67Fb	chitin binding	yes	Findlay, Takemori
CG1837	X	pretaporter	prtp	Thioredoxin		Takemori
CG18495	2R	Proteasome alpha1 subunit	Prosalpha1	Proteasome α1 subunit		Takemori
CG18628	3L	-	CG18628	unknown	yes	Findlay
CG18749	3R	-	CG18749	Collagen like		Wolfner
CG2331	2R	TER94	TER94	Atpase, serine protease, cdc48		Takemori
CG2341	3R	Ccp84Ad	Ccp84Ad	cuticle binding	yes	Findlay
CG2665	2R	Protein ejaculatory bulb II	PebII	unknown	yes	Findlay
CG2668	2R	Protein ejaculatory bulb	Peb	unknown	yes	Findlay
CG2767	3R	-	CG2767	oxidoreductase		Takemori
CG2826	2L	lectin-21Ca	lectin-21Ca	C-type lectin		Wolfner
CG2852	2R	-	CG2852	protein folding	yes	Findlay, Wolfner
CG2918	X	-	CG2918	chaperone		Takemori, Wolfner
CG2975	2L	-	CG2975	peptide/Prohormone		Takemori, Wolfner

CG	LOCATION	NAME	SYMBOL	FUNCTION	TRANSFER	LIST OVERLAP
CG3011	X	-	CG3011	Serine hydroxymethyltransferase		Takemori
CG30395	2R	-	CG30395	peptide/Prohormone	yes	Findlay, Wolfner
CG30448	2R	Odorant-binding protein 56i	Obp56i	Oderant Binding	yes	Findlay
CG30450	2R	Odorant-binding protein 56f	Obp56f	Oderant Binding	yes	Findlay
CG30463	2R	-	CG30463	lectin		Takemori, Wolfner
CG30473	2R	Odorant-binding protein 51a	Obp51a	Oderant Binding	yes	Findlay
CG30486	2R	-	CG30486	CRISP	yes	Findlay, Wolfner
CG30488	2R	Antare	CG30488	CRISP	yes	Findlay, Wolfner
CG3074	2R	Secreted Wg-interacting molecule	Swim	cysteine protease		Takemori
CG31016	3R	-	CG31016	Collagen like		Takemori, Wolfner
CG31021	3R	-	CG31021	Hydroxylase		Takemori
CG31056	3R	Accessory gland protein 98AB	Acp98AB	peptide/Prohormone		Wolfner
CG31418	3R	-	CG31418	peptide/Prohormone	yes	Findlay, Wolfner
CG31419	3R	-	CG31419	unknown	yes	Findlay
CG31472	3R	-	CG31472	phosphate oxidase		Takemori
CG31515	3R	-	CG31515	serpin	yes	Findlay, Takemori, Wolfner
CG31519	3R	Odorant receptor 82a	Or82a	Oderant Binding	yes	Findlay
CG3153	3R	Niemann-Pick type C-2b	Npc2b	sterol binding	yes	Findlay
CG31659	2L	-	CG31659	peptide/Prohormone	yes	Findlay, Wolfner
CG31680	2L	-	CG31680	peptide/Prohormone	yes	Findlay, Wolfner
CG31704	2L	-	CG31704	serpin	yes	Findlay, Takemori
CG31758	2L	-	CG44008	protease inhibitor	yes	Findlay
CG31779	2L	Acp24A4	Acp24A4	protease inhibitor	yes	Findlay
CG31872	2L	-	CG31872	acid lipase	yes	Findlay, Wolfner
CG31883	2L	-	CG31883	unknown	yes	Findlay, Takemori
CG31941	2L	Odorant-binding protein 22a	Obp22a	Oderant Binding	yes	Findlay, Takemori
CG3210	2L	Dynamin related protein 1	Drp1	Gtpase		Takemori
CG32190	3L	NUCB1	NUCB1	DNA binding	yes	Findlay
CG32197	3L	-	Met75Ca	unknown	yes	Findlay
CG32201	3L	-	CG32201	Hydroxylase		Wolfner
CG32203	3L	Serpin 75F	Spn75F	serpin	yes	Findlay, Wolfner
CG32382	3L	sphinx2	sphinx2	serine protease		Takemori, Wolfner
CG32383	3L	sphinx1	sphinx1	serine protease		Wolfner

CG	LOCATION	NAME	SYMBOL	FUNCTION	TRANSFER	LIST OVERLAP
CG32667	X	-	CG32667	Collagen like	yes	Findlay, Wolfner
CG32833	2R	-	CG32833	serine protease	yes	Findlay, Wolfner
CG33126	2L	Neural Lazarillo	NLaz	lipid binding	yes	Findlay
CG3322	3L	Laminin B2	LanB2	Laminin		Takemori
CG33259	3L	-	CG33259	peptide/Prohormone	yes	Findlay, Wolfner
CG33462	2R	-	CG33462	serine protease		Wolfner
CG3349 .	3L	-	CG3349	peptide/Prohormone	yes	Findlay, Wolfner
CG33495	3R	Ductus ejaculatorius peptide 99B	Dup99B	peptide/Prohormone	yes	Findlay
CG33530	2R	Accessory gland protein 53C14c	Acp53C14c	unknown	yes	Findlay
CG3359	3R	midline fasciclin	mfas	fasciclin	yes	Findlay, Wolfner
CG33943	3L	BG642312	BG642312	peptide/Prohormone	Yakuba, Simulans	Findlay, Takemori, Wolfner
CG34033	2R	-	CG34033	C-type lectin	yes	Findlay, Takemori
CG34034	3R	-	CG34034	unknown	yes	Findlay
CG34051	2L	-	CG34051	peptide/Prohormone		Wolfner
CG34053	3R	-	CG42824	peptide/Prohormone		Takemori, Wolfner
CG34102	2L	BG642163	BG642163	peptide/Prohormone		Wolfner
CG34103	3R	-	BG642167	peptide/Prohormone		Takemori, Wolfner
CG34130	3R	-	CG34130	serine protease	yes	Findlay
CG34435	X	-	CG34435	Atpase	yes	Findlay, Takemori
CG3640	2R	-	CG3640	CRISP		Wolfner
CG3801	3L	Accessory gland protein 76A	Acp76A	serpin	yes	Findlay, Wolfner
CG3832	2R	Peptidylglycine -alpha-hydroxylating monooxygenase	Phm	peptidyglycine	yes	Findlay, Takemori
CG3937	3R	cheerio	cher	Filamin		Takemori
CG4147	X	Heat shock 70-kDa protein cognate 3	Hsc70-3	chaperone		Wolfner
CG4605	2L	Accessory gland protein 32CD	Acp32CD	peptide/Prohormone		Takemori, Wolfner
CG4634	2R	Nucleosome remodeling factor - 38kD	Nurf-38	pyrophosphatase		Takemori
CG4815	3R	-	CG4815	serine protease		Wolfner
CG4847	2R	-	CG4847	cysteine protease		Wolfner
CG4986	3R	Male-specific transcript 57Dc	Mst57Dc	peptide/Prohormone		Takemori, Wolfner
CG5016	3R	Male-specific RNA 57Db	Mst57Db	peptide/Prohormone		Wolfner
CG5119	2R	-	pAbp	mRNA binding		Takemori
CG5162	X	-	CG5162	lipase	yes	Findlay

CG	LOCATION	NAME	SYMBOL	FUNCTION	TRANSFER	LIST OVERLAP
CG5177	2L	-	CG5177	phosphatase		Takemori
CG5207	3R	SCP-containing protein A	scpr-A	unknown	yes	Findlay
CG5267	2R	-	CG5267	trypsin protease inhibitor	yes	Findlay
CG5402	3R	-	CG5402	unknown	yes	Findlay
CG5450	2L	Cytoplasmic dynein light chain 2	Cdlc2	ATPase activity	yes	Findlay
CG5520	3R	Glycoprotein 93	Gp93	chaperone		Wolfner
CG5654	3L	ypsilon schachtel	yps	DNA binding		Takemori
CG5809	2L	calcium-binding protein 1	CaBP1	isomerase activity		Takemori
CG6069	3R	-	CG34129	serine protease		Wolfner, Takemori
CG6071	3L	-	CG6071	metallo protease	yes	Findlay
CG6168	3L	-	CG6168	serine protease		Wolfner
CG6289	3L	Serpin 77Bc	Spn77Bc	serpin	yes	Findlay, Wolfner
CG6352	X	Ods-site homeobox	OdsH	homeobox	yes	Findlay
CG6426	2R	-	CG6426	lysozyme	yes	Findlay
CG6461	X	gamma-glutamyl transpeptidase	Ggt-1	peptidase	yes	Findlay, Wolfner
CG6555	2L	-	CG6555	peptide/Prohormone	yes	Findlay, Wolfner, Takemori
CG6663	3L	Serpin 77Bb	Spn77Bb	serpin	yes	Findlay
CG6690	3R	-	CG6690	Thioredoxin	yes	Findlay
CG6717	2L	Serpin 28B	Spn28B	serpin	yes	Findlay
CG6917	3L	Esterase 6	Est-6	carboxylesterase	yes	Findlay
CG6988	3L	Protein disulfide isomerase	Pdi	protein folding		Wolfner, Takemori
CG7157	2L	Accessory gland protein 36DE	Acp36DE	Glycoprotein	yes	Findlay, Wolfner, Takemori
CG7225	2R	windbeutel	wbl	protein binding		Takemori
CG7304	3L	-	CG7304	peptide/Prohormone		Wolfner, Takemori
CG8050	3R	Cystatin-like	Cys	protease inhibitor	yes	Findlay
CG8093	2R	-	CG8093	acid lipase		Wolfner
CG8102	2R	-	CG8102	dehydrogenase	yes	Findlay
CG8137	2L	Serpin 28F	Spn28F	serpin	yes	Findlay, Wolfner
CG8194	3L	Ribonuclease X25	RNaseX25	RNase		Takemori, Wolfner
CG8286	3R	P58IPK	P58IPK	Tetratricopeptide		Takemori
CG8420	3R	-	CG8420	peptide/Prohormone		Wolfner
CG8462	2R	Odorant-binding protein 56e	Obp56e	Oderant Binding	yes	Findlay
CG8622	2R	Accessory gland protein 53Ea	Acp53Ea	peptide/Prohormone	yes	Findlay, Wolfner

CG	LOCATION	NAME	SYMBOL	FUNCTION	TRANSFER	LIST OVERLAP
CG8626	2R	Accessory gland protein 53C14a	Acp53C14a	peptide/Prohormone	yes	Findlay, Wolfner
CG8651	3R	trithorax	trx	methyltransferase	yes	Findlay
CG8982	2L	Accessory gland protein 26Aa	Acp26Aa	peptide/Prohormone	yes	Findlay, Wolfner
CG9006	2R	Enigma	Egm	Acyl-CoA dehydrogenase, type 1	yes	Findlay, Takemori
CG9024	2L	Accessory gland protein 26Ab	Acp26Ab	peptide/Prohormone	yes	Findlay, Wolfner
CG9029	2L	-	CG9029	peptide/Prohormone	yes	Findlay, Wolfner
CG9074	3R	Male-specific RNA 57Da	Mst57Da	peptide/Prohormone	yes	Findlay, Wolfner
CG9111	3L	Lysozyme C	LysC	lysozyme	yes	Findlay
CG9168	3L	-	CG9168	phosphatase	yes	Findlay
CG9334	2L	Serpin 38F	Spn38F	serpin	yes	Findlay, Wolfner
CG9359	3R	beta-Tubulin at 85D	betaTub85D	GTPase	yes	Findlay, Takemori
CG9429	3R	Calreticulin	Crc	chaperone		Wolfner
CG9456	2R	Serpin 42Dd	Spn42Dd	serpin	yes	Findlay, Takemori
CG9525	2L	-	CG9525	peptide/Prohormone		Takemori, Wolfner
CG9748	3R	belle	bel	helicase		Takemori
CG9806	X	-	CG9806	Aminopeptidase		Wolfner
CG9847	2R	FK506-binding protein 14 ortholog	Fkbp14	protein folding		Wolfner
CG9975	2R	-	CG9975	unknown	yes	Findlay, Takemori
CG9997	3R	-	CG9997	serine protease	yes	Findlay, Wolfner

APPENDIX C

A SCREEN OF HIGHLY ABUNDANT ACCESSORY GLAND PROTEINS REVEALS FOUR NEW PLAYERS IN REGULATING THE FEMALE POST-MATING RESPONSE

The majority of the data in this chapter is preliminary, was collected early in my tenure as a graduate student, and is included here to help future students. The genes I screened were originally chosen by a previous post doc in the lab, K. Ravi Ram, from a list of genes he complied in 2007 [1]. All of the genes listed in Table C.1 are highly expressed in the male accessory gland [2] and some (CG12558 for example) are in the top 50 transcripts detected. Since CG12558 (*intrepid*) is the subject of Chapter 4 only supplemental information about CG12558 is presented here.

To test the function of these genes in regulating the PMR I conducted fertility fecundity assays similar to those described in previous chapters. To generate knockdown males I crossed UAS-RNAi lines (listed in Table C.1) to *tubulin*-GAL4. In the case of CG32383 and CG30463 ubiquitous knockdown with *tubulin*-GAL4 was lethal. To avoid lethality I used *prd*-GAL4 for these lines instead. Unfortunately, there was an issue with the *prd*-GAL4 driver and the VDRC background such that all males generated with this driver displayed LTR phenotypes (Figure C.1) regardless of the gene knocked down. As such, *prd*-GAL4 data has been omitted. Knockdown was check by RT-PCR, the results of which are summarized in Table C.1. Because two of the candidate lines, CG32201 and CG18234, did not knock down they are also not included in the data. Virgin Canton-S females were then mated to either a control (balancer

Table C.1: Candidate Genes

CG#	Function	Signal sequence	Location	RNAi line	Knockdown	Other available lines
CG7304	Glycosyl transferase	no	3L	34259	yes	yes
CG15116	Glutathione peroxidase	no	2R	30877	yes	yes
CG18233	oxidoreductase activity	yes	3L	40141	yes	yes
CG31704	Kazal serine protease	yes	2L	39451	yes	yes
CG18234	oxidoreductase activity	yes	3L	19187	no	no
CG32201	deoxygenase activity	yes	3L	47008	no	yes
CG32383	protease	yes	3L	43241	yes	yes
CG32667	unknown	yes	X	49039	yes	yes
CG31016	deoxygenase activity	yes	3R	21280	yes	no
CG30463	GalNAc transferase	yes	2R	4924	yes	yes
CG10041	serine protease	yes	3R	47640	yes	yes

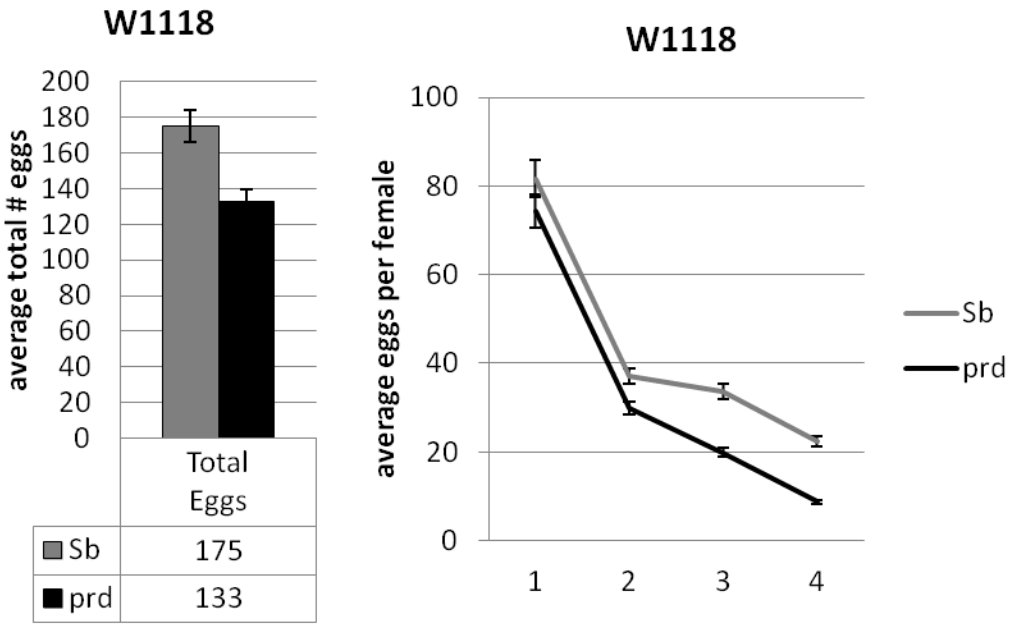


Figure C.1: The *prd*-GAL4 driver line produces an LTR effect on egg-laying

An egg-laying assay comparing siblings that carry the *prd*-GAL4 driver to their stubble (Sb) balancer siblings. Mates of males carrying the *prd*-GAL driver lay significantly fewer eggs (rmANOVA: p=<0.05, Sb) in 4 days after mating than mates of control (Sb) males. Similar LTR like effects were seen for receptivity and SP storage (not shown). These results suggest that *prd*-GAL4 is not an appropriate driver line for studying the LTR.

sibling) or RNAi male. None of the males knocked down for candidate genes influenced egg-laying in mated females (Figure C.2A). However, mates of CG31704 males showed a significant decrease in hatchability (#progeny/#eggs) (Figure C.2B). A summary of the outcomes of the assays is in Table C.2. These results suggest that none of these six genes that were successfully tested are essential for egg-laying and as such are likely not important for the LTR.

To verify that these genes were not important for the LTR, I performed receptivity assays as described in previous chapters. This time, *Hsp70*-GAL4 was used instead of *prd*-GAL4 to knockdown CG30463 and CG32383. Males were raised at room temperature and then heat-shocked as described in Chapter 2. All RNAi males were able to suppress remating in their mates at 1d after the start of mating (ASM) (Figure C.3, summarized in Table C.2). However, mates of CG31704 and CG32383 males were more receptive than mates of control males at 4d ASM. No differences were seen at 10d ASM for any of the RNAi males. These results suggest that CG31704 and CG32383 are important for regulating female receptivity. Reduced expression of CG31704 does not impact egg-laying, suggesting that these two phenotypes may be separable. The influence of CG32383 on female egg-laying is currently unknown. While these genes were tested for their impact on SP storage, that work was done using *prd*-GAL4 and as such is not informative.

To confirm the receptivity results and attempt to identify the cellular source of these genes I used tissue specific drivers *iab-6D1*-GAL4 (Chapter 6) and *Acp26Aa*-GAL4 [3] to knock down CG31704, CG32383, CG12558, and CG30463 in the accessory gland. I included CG30463 despite observing negative results using heat-shock because heat-shock depends on fast turnover of proteins and sometimes returns false negative results. Before testing, I confirmed that both drivers are expressed normally in the accessory glands by crossing each driver to UAS-

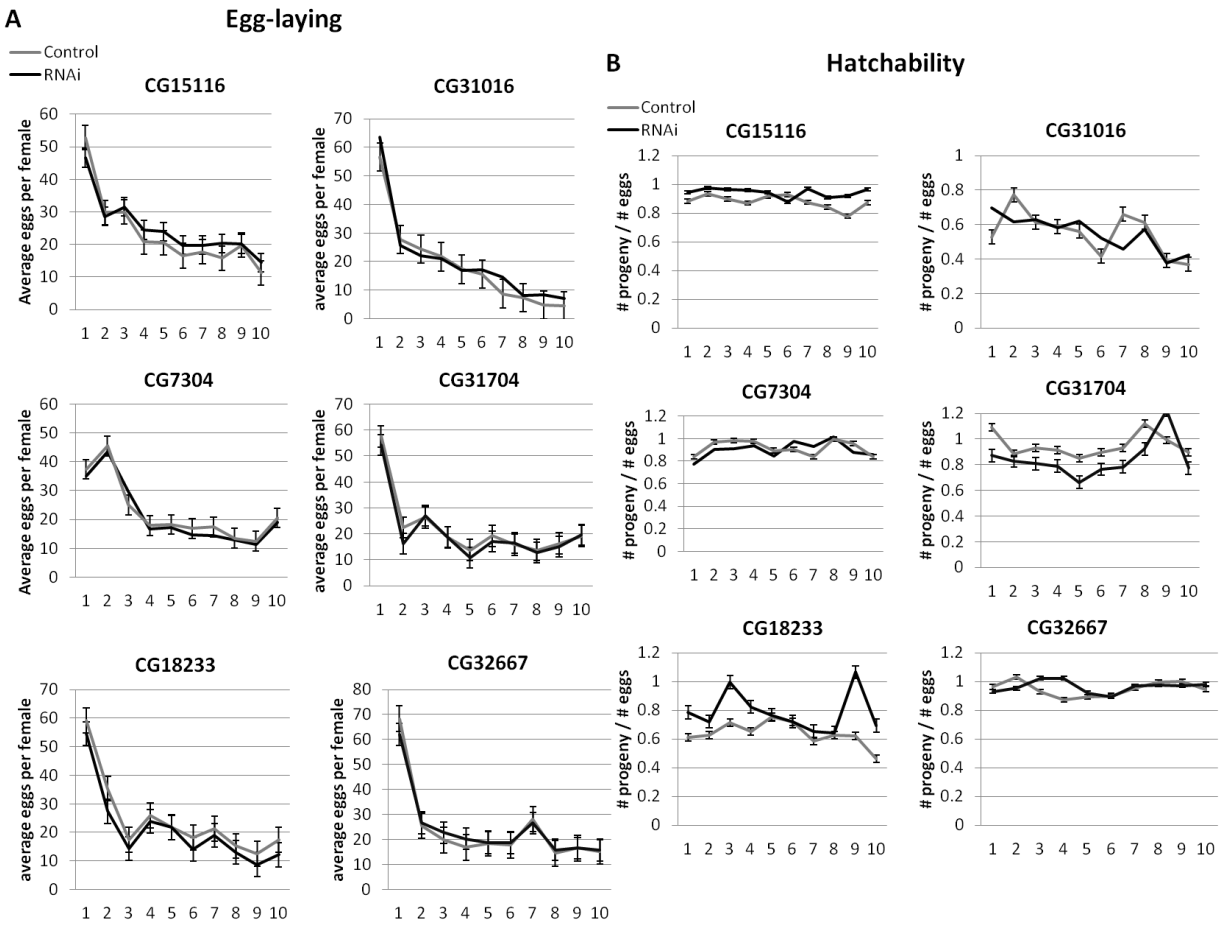


Figure C.2: Egg-laying and hatchability in mates of RNAi males over 10 days

Virgin females were mated to either control (balancer sibling) or RNAi males for each of 6 genes, CG15116, CG31016, CG7304, CG31704, CG18233, and CG32667. No differences were seen in egg-laying for any of the trials. Mates of CG31704 males have decreased progeny production, resulting in a decrease in hatchability (# progeny/# eggs laid) over the first 8 days after mating (rmANOVA p=0.010). Summary statistics and Ns are in Table C.2.

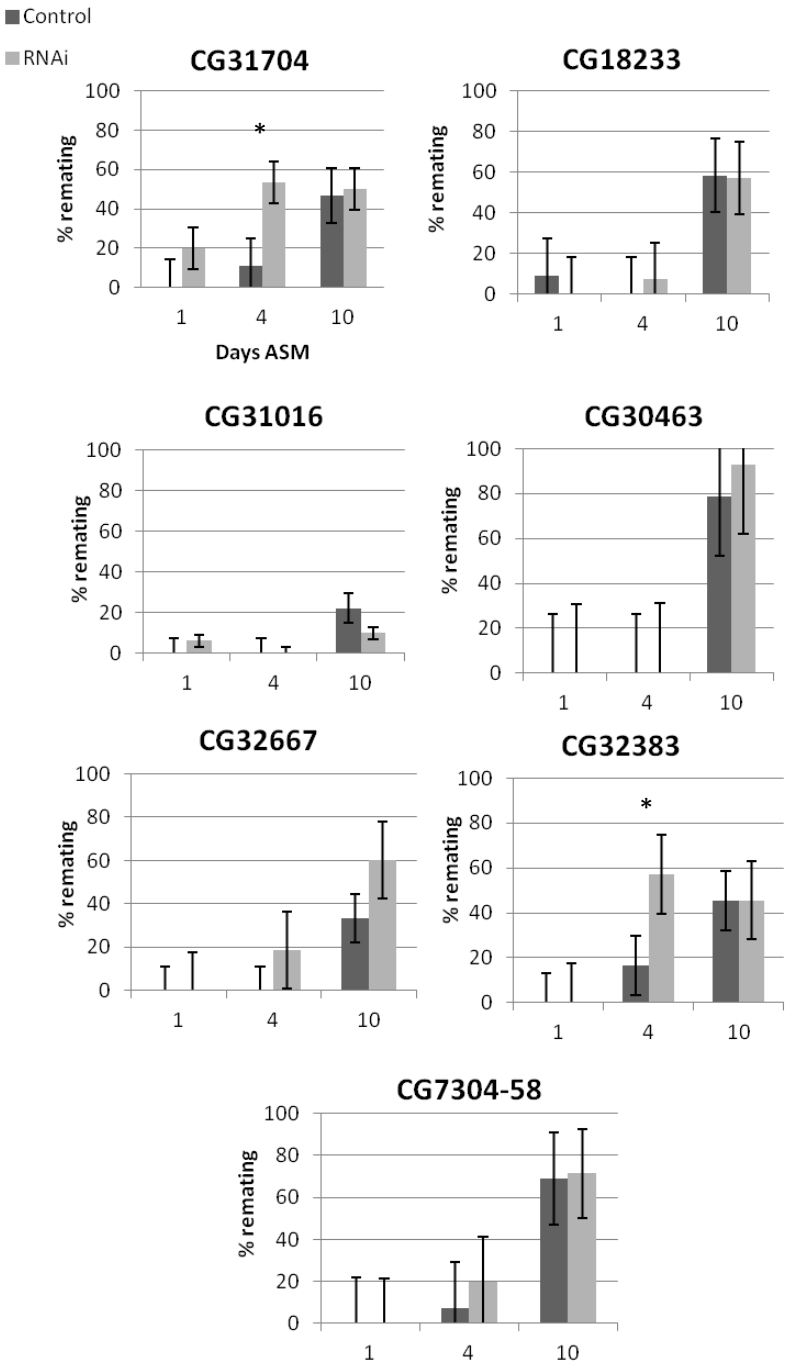


Figure C.3: Receptivity in mates of RNAi males

Virgin Canton-S females were mated to either control (balancer sibling) or RNAi males, isolated, and then allowed access to a single Canton-S male at 1d, 4d, or 10d ASM. Both CG32383 (WRST p=0.038) and CG31704 (WRST p=<0.001) influenced female receptivity at 4d ASM. A summary of statistics and Ns is in Table C.2.

Table C.2: Statistics and sample sizes for fertility/fecundity and receptivity assays

					Receptivity			
CG#	**FFA**	**N**	**Hatchability**	**driver**	**1d**	**4d**	**10d**	**Ns**
CG7304	0.895	12-18	0.496	*tubulin*	n.s	n.s	n.s	10-20
CG15116	0.579	13-17	0.417	*tubulin*	-	-	n.s	11-13
CG18233	0.323	12-13	0.154	*tubulin*	n.s	n.s	n.s	11-14
CG31704	0.675	13-14	0.299 (0.010 for days 1-8)	*tubulin*	n.s	<0.001	n.s	10-30
CG18234	-	-	-	*tubulin*	-	-	-	-
CG32201	-	-	-	*tubulin*	-	-	-	-
CG32383	prd	-	-	*Hsp70*	n.s	0.038	n.s	11-14
CG32667	0.832	15-18	0.608	*tubulin*	n.s	n.s	n.s	9-16
CG31016	0.550	14-13	0.987	*tubulin*	n.s	n.s	n.s	9-16
CG30463	prd	-	-	*Hsp70*	n.s	n.s	n.s	11-14
CG10041	prd	-	-	*tubulin*	n.s	n.s	n.s	10-16

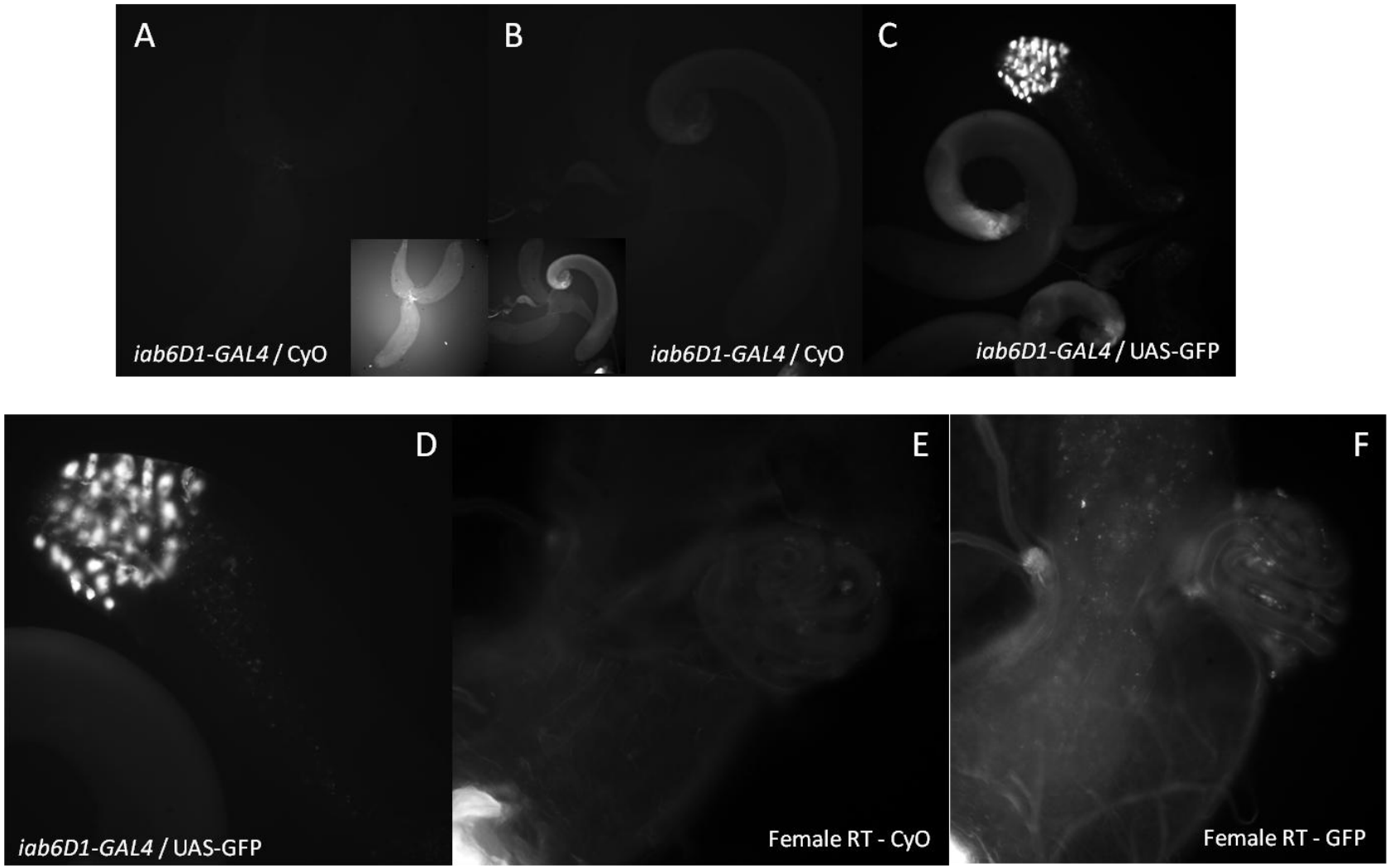
rmANOVA was used to assess FFA and Hatching data whereas WRST was used to assess receptivity data. All analysis was performed using jmp10 software.

mCD8GFP, a membrane bound GFP (Figure C.4&5). Note that GFP is not detectable in most of the secondary cells of *Acp26Aa*-GAL4 males (Figure C.5). For controls, sibling controls were used in the iab-6D1-GAL4 crosses whereas a driver control was used for *Acp26Aa*-GAL4 (this stock is not balanced, the driver is on the X chromosome). In all cases, knockdown with *iab-6D1*-GAL4 resulted in a significant increase in female receptivity at 4d ASM (for evidence that the driver does not always produce a positive result, unlike *prd*-GAL4 see Chapter 6) whereas knockdown with *Acp26Aa*-GAL4 did not (CG32383 not tested) (Figure C.6). These results suggest that all 4 genes are necessary in the secondary cells for regulating female receptivity at 4d ASM and also that our *Hsp70*-GAL4 protocol works for detecting this in CG32383 but not for CG30463. Egg-laying assays for CG32383 and CG30463 using the *iab-6D1*-GAL4 driver are needed to determine whether these genes confer a standard LTR phenotype (both egg-laying and receptivity affected) or if they behave like CG31704 and only affect receptivity.

Together these results suggest that the Sfps CG12558, CG31704, CG32383, and CG30463 are expressed in the secondary cells. Further, that their expression in the secondary

Figure C.4: Expression of *iab-6D1*-GAL4

Male reproductive tracts and lower female reproductive tracts from *iab-6D1*-GAL4 males or sibling controls (CyO) expressing membrane bound mCD8;;GFP and their mates. Glands were dissected into 1xPBS, not fixed, and immediately observed using a Leica CTR5000 microscope **A&B)** Control males do not show auto-fluorescence in the accessory gland. The insets are overexposed to show the location of the reproductive tract in the figure. **C&D)** Males carrying the *iab-6D1*-GAL4 driver express GFP strongly in their secondary cells. There is some auto-fluorescence in the testis and potentially some GFP expression near the entry into the seminal vesicle, however this was variable and not quantified. **D)** A clearer view of the secondary cells enlarged from panel C shows the presence of GFP in the lumen of the accessory gland in distinct puncta. **E)** Mates of control males do not have these GFP puncta in their uterus. **F)** Mates of iab-6D1-GFP males do show GFP puncta in their uterus. Whether these puncta are membranous vesicles, pieces of broken membrane, or an artifact is unclear at this time.


A
B
C
iab6D1-GAL4 / CyO
iab6D1-GAL4 / CyO
iab6D1-GAL4 / UAS-GFP
D
E
F
iab6D1-GAL4 / UAS-GFP
Female RT - CyO
Female RT - GFP

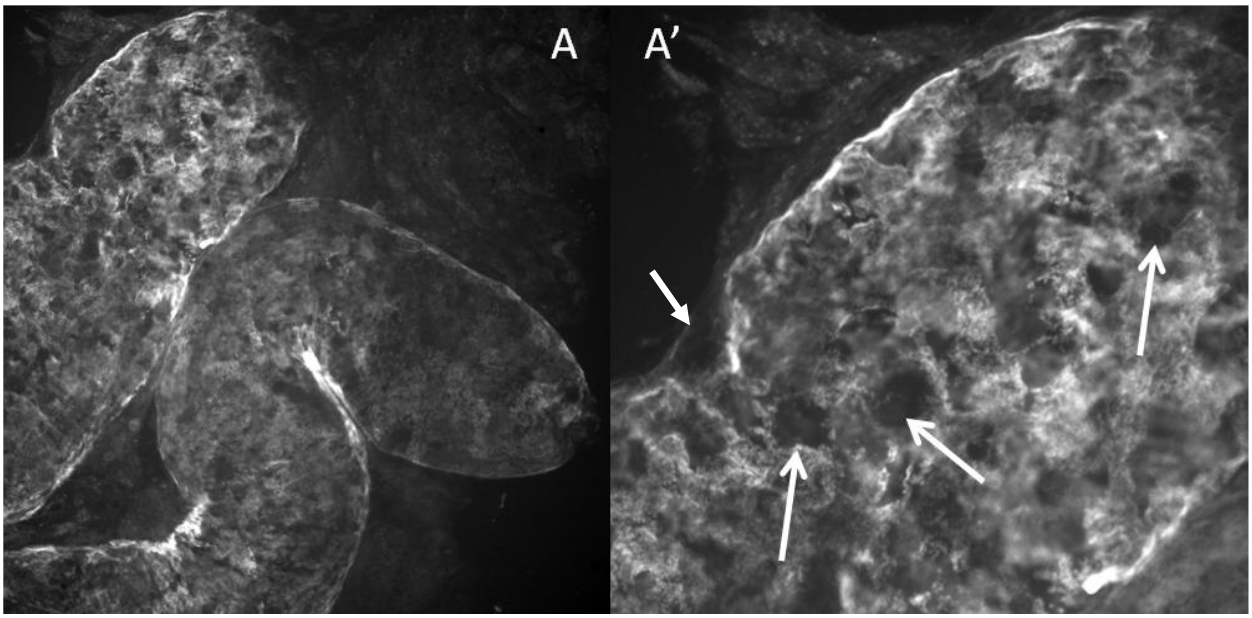


Figure C.5: GFP expression in Acp26Aa-GAL4 males is not in the secondary cells at 4 days post eclosion

Ovulin proteins is detectable inside both cell types of the accessory gland in young males but only in the vacuole of the secondary cells in older unmated males [4]. Despite this, we fail to see GFP in the secondary cells of males carrying the Acp26Aa-GAL4 driver at the ages used during our standard fertility/fecundity and receptivity assays. **A)** Accessory glands from *Acp26Aa*-GAL4;UAS-*mCD8*;;GFP males. **A')** Enlargements of A&B with arrows pointing at some of the secondary cells that are not expressing GFP.

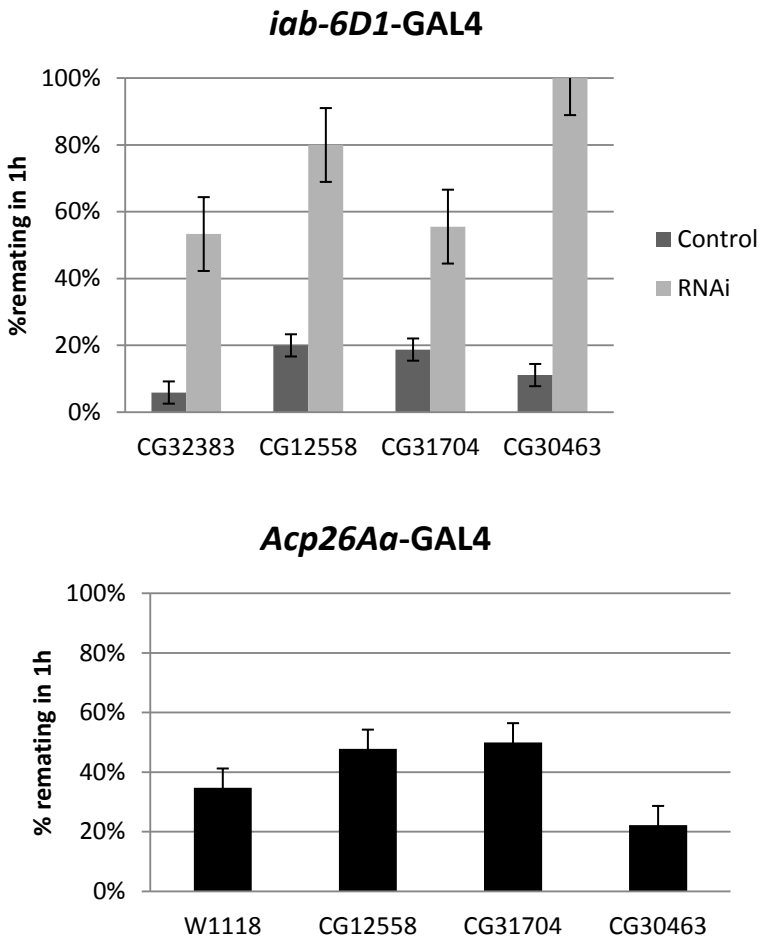


Figure C.6: 4d receptivity *in iab-6D1*-GAL4 and *Acp26Aa*-GAL4 RNAi males

Knockdown of CG32383, CG12558, CG31704, and CG30463 in the secondary cells resulted in increased receptivity compared to mates of control males (WRST all p=<0.001, Ns 15-22). Conversely, knockdown of CG12558, CG31704, and CG30463 using Acp26Aa-GAL4, which expressed poorly if at all in the secondary cells of males at the age used for these assays (3-5 days old), did not (CG12558 WRST p=0.550 N=23, CG31704 WRST p=0.3482 N=16, CG30463 WRST p=0.3862 N=18, W1118 N=23). This suggests that all four of these genes are essential in the secondary cells for long term receptivity and at least three are not needed in the main cells.

cells is essential for their influence over 4d receptivity in their mates. Due to issues with the *prd*-GAL4 driver and time constraints while pursuing the work in the body of this thesis it is still unclear whether 3 of these genes (CG31704, CG32383, and CG30463) are important for the storage of SP. Likewise, no attempt has been made to place these genes into the LTR network. Further, the possibility that one (CG31704) or more of these genes impact receptivity but not egg-laying is curious and speaks to the prospect that these two usually concurrent LTR effects might be separable. The opposite effect was seen in mates of CG9036 RNAi males (Chapter 6), which affected egg-laying but not receptivity. It is possible that these anomalous observations are dosage based due to the constraints of RNAi knockdown, but even this may be useful for teasing apart the branches of the LTR network.

REFERENCES

1. Ram KR, Wolfner MF (2007) Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. Integrative & Comparative Biology 47: 19p.
2. Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39: 715-720.
3. Chapman T, Bangham J, Vinti G, Seifried B, Lung O, et al. (2003) The sex peptide of Drosophila melanogaster: female post-mating responses analyzed by using RNA interference. Proc Natl Acad Sci U S A 100: 9923-9928.
4. Monsma SA, Harada HA, Wolfner MF (1990) Synthesis of two Drosophila male accessory gland proteins and their fate after transfer to the female during mating. Dev Biol 142: 465-475.